

Walker, D.H.
 Rickettsia felis: molecular characterization of a new member of the
 spotted fever group
 Int. J. Syst. Evol. Microbiol. 51 (Pt 2), 339-347 (2001).
 MEDLINE
 21217364
 PUBMED
 11321078
 2 (bases 1 to 5513)
 Bouyer, D.H., Stenos, J., Crocquet-Valdes, P.A., Foil, L.D. and
 Walker, D.H.
 Direct Submission
 Submitted (30-SEP-1999), Pathology, University of Texas Medical
 Branch at Galveston, 301 University Blvd., Galveston, TX
 77555-0609, USA

FEATURES

source
 1. 5513
 /organism="Rickettsia felis"
 /db_xref="taxon:42862"
 478..2337
 /gene="ompA"
 478..2337
 /gene="ompA"
 /note="truncated as compared to other Rickettsia ompA
 proteins"
 /codon_start=1
 /transl_table=11
 /product="outer membrane protein A"
 /protein_id="AAG17185.1"
 /db_xref="GI:10441769"
 /translation="MANISLKFQKIQKLTALFTTSTAAIMLGGSGVLGAARTVT
 ADCAELAAAGTNGICPGAGFVAGSTLOYTGAFVTGADVSRLDILNFAAGLESVTGD
 ISLGSVVDVGGKLANVDDGLTLTGTAAYGANPALLFQGGGAANNRYTALIG
 NYTLGGANAGLTASDPVLGPILAGINDGGIITDNDAAINGTNTNPAQISL
 GASTLGGGAVIKATTKLTNAAPVLTNNANAVLTCAVDNMTGGDDVGVNLNGALS
 QVTGNTGNTSLATISVGATGLGAVIKATTKLTNAASVLTNAVLNATVLDVNT
 GDNVGVNLNGALSQVNTGNTSLATINIGAVTLDGAVIKATTKLTDDASVL
 IFNPVVTGAIDONTNANKGVVIFTGASTVDNIGNTAVLAEVSGAGLQIQGGV
 KANALNTONASVVTGDSVTGSGIGTELFATVIGAITLRAGSLAANNIDFGA
 ASNLEPGRKNNYLIGTIANGLNATVANDVSIGTVAQINQNNKIFV
 INAKNDVLDQAQISFKAARSLFANVLSQIMLSLKIIPVLLTVVVS"
 BASE COUNT 1814 a 961 c 1181 g 1557 t
 ORIGIN
 Query Match 100.0%; Score 1920; DB 1; Length 5513;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGCGGAATATTTCTTAAATATTTCAAAAAGCAATTCAAAAGGCTTTAAAACGTCT 60
 DB 478 ATGCGGAATATTTCTTAAATATTTCAAAAAGCAATTCAAAAGGCTTTAAAACGTCT 537
 QY 61 TTATTCACCACTCAACCGCAGCGATATGCTAACCGGTAGTGAGTCCCTGGTGGCA 120
 DB 538 TTATTCACCACTCAACCGCAGCGATATGCTAACCGGTAGTGAGTCCCTGGTGGCA 597
 QY 121 AGAACCTGAAGTCTGATGTCAGAGCTTGCACCGGGAACAATATAGTCCCTGGAGCC 180
 DB 598 AGAACCTGAAGTCTGATGTCAGAGCTTGCACCGGGAACAATATAGTCCCTGGAGCC 657
 QY 181 GGTGCTTTTGTAGCGGGTCTTACATTTACAATATACCGGTGCTTTTACGGTTACTGATGCT 240
 DB 658 GGTGCTTTTGTAGCGGGTCTTACATTTACAATATACCGGTGCTTTTACGGTTACTGATGCT 717
 QY 241 GACGTAAGTGTCTGTCATTAGATTTAAATAATTTTCAGCAGGCTCTTTTTCAGTAAC 300
 DB 718 GACGTAAGTGTCTGTCATTAGATTTAAATAATTTTCAGCAGGCTCTTTTTCAGTAAC 777
 QY 301 GGTGATATTCATAGTCTCAGTGTAGATACCGGAGGAGCTAATAACCTTGCAGTTAAT 360
 DB 778 GGTGATATTCATAGTCTCAGTGTAGATACCGGAGGAGCTAATAACCTTGCAGTTAAT 837
 QY 361 ATTGATGATGTTTAACTTAACTTTAACAGGTACCGGCTACTGAGCGCTACGGTGCAGAA 420
 DB 838 ATTGATGATGTTTAACTTAACTTTAACAGGTACCGGCTACTGAGCGCTACGGTGCAGAA 897

QY 421 CTCCGGTGTATTCCAAAGTGGACAAGCTGCTCTAATAATACATATACCTGCTTTAGT 480
 DB 898 CTCCGGTGTATTCCAAAGTGGACAAGCTGCTCTAATAATACATATACCTGCTTTAGT 957
 QY 481 AATATACTCTAGTGGAGGAATGCGGTTTGACTATTGCTTCAGATCCAGATGTATTA 540
 DB 958 AATATACTCTAGTGGAGGAATGCGGTTTGACTATTGCTTCAGATCCAGATGTATTA 1017
 QY 541 GGACCAATAACGCTTGCAGGAATAATAGATGGAGAGGTATAAATACCTGACAATACAGAT 600
 DB 1018 GGACCAATAACGCTTGCAGGAATAATAGATGGAGAGGTATAAATACCTGACAATACAGAT 1077
 QY 601 GCTGCCATTAACGGAACAATAAGTAACTAATCCGCGAGCTCAAAATAAGCATTTGGAGCA 660
 DB 1078 GCTGCCATTAACGGAACAATAAGTAACTAATCCGCGAGCTCAAAATAAGCATTTGGAGCA 1137
 QY 661 AGTACGCTTCTCTTGGAGGGGAGTATTAAAGCCACTAGGACTAAGTTAAACAATGCA 720
 DB 1138 AGTACGCTTCTCTTGGAGGGGAGTATTAAAGCCACTAGGACTAAGTTAAACAATGCA 1197
 QY 721 GCGCGGTATTAACACTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 780
 DB 1198 GCGCGGTATTAACACTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1257
 QY 781 ACAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 DB 1258 ACAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1317
 QY 841 AATATAGTATACAAATTCATTAGCCACCAATAGTGTAGGAGCAGGTACGGCTACGTTA 900
 DB 1318 AATATAGTATACAAATTCATTAGCCACCAATAGTGTAGGAGCAGGTACGGCTACGTTA 1377
 QY 901 GGGGAGCGGTATTAAAGCCACCTACCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 960
 DB 1378 GGGGAGCGGTATTAAAGCCACCTACCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1437
 QY 961 CTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1020
 DB 1438 CTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1497
 QY 1021 GTCGTAATCTAAGTGGAGCAATTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1080
 DB 1498 GTCGTAATCTAAGTGGAGCAATTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1557
 QY 1081 TTAGCCACGATAAATATAGGAGCAGGTGCTGCTACCTTGGATGGAGCGGTATTAAAAGCT 1140
 DB 1558 TTAGCCACGATAAATATAGGAGCAGGTGCTGCTACCTTGGATGGAGCGGTATTAAAAGCT 1617
 QY 1141 ACTACGACTAAGTTAACAGATGATGCTGCTAGTATTAATATTTACGAATCCTGTAGTAGTA 1200
 DB 1618 ACTACGACTAAGTTAACAGATGATGCTGCTAGTATTAATATTTACGAATCCTGTAGTAGTA 1677
 QY 1201 ACCGAGCAATAGATAATACCGGTAATGCCAATAAAGGTGCTGCTTTACCGGAGCA 1260
 DB 1678 ACCGAGCAATAGATAATACCGGTAATGCCAATAAAGGTGCTGCTTTACCGGAGCA 1737
 QY 1261 AGTACGCTAACCGGTAATATAGTAAACACCGCAGTATTAGCAGAGGTAAAGGTAGGAGCA 1320
 DB 1738 AGTACGCTAACCGGTAATATAGTAAACACCGCAGTATTAGCAGAGGTAAAGGTAGGAGCA 1797
 QY 1321 GGTGCTGCAAAATACAAAGCGGAGTAGTAAAAGCGAATGCAATAAAGTAACTTAAACGGATAAT 1380
 DB 1798 GGTGCTGCAAAATACAAAGCGGAGTAGTAAAAGCGAATGCAATAAAGTAACTTAAACGGATAAT 1857
 QY 1381 GCGTCAGTACTTACATTTACCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 DB 1858 GCGTCAGTACTTACATTTACCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1917
 QY 1441 CTCTTCGCAACAGTGAATATAGGAGCAGGAATAACATTTACGAGCGGAGGAGGAGGAGGAGGAGGAGGAG 1500
 DB 1918 CTCTTCGCAACAGTGAATATAGGAGCAGGAATAACATTTACGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1977

QY 1501 GCGAATAATATAGATTTTGGAGCTCGAGTAAATTTAGAGTTTAAACGGTCTCCGCGTAAG 1560
 |||||||
 Db 1978 GCGAATAATATAGATTTTGGAGCTCGAGTAAATTTAGAGTTTAAACGGTCTCCGCGTAAG 2037
 |||||||
 QY 1561 AATTAATACTTAATCGGAACCTATAGCAACGGTATATATAGTCTACTTAAATATTAAATGCT 1620
 |||||||
 Db 2038 AATTAATACTTAATCGGAACCTATAGCAACGGTATATATAGTCTACTTAAATATTAAATGCT 2097
 |||||||
 QY 1621 GCTGTACAGTGATTTGCAAAATGATCTTAGTAGTACAGCTTGCCAAATTAACATTCAA 1680
 |||||||
 Db 2098 GCTGTACAGTGATTTGCAAAATGATCTTAGTAGTACAGCTTGCCAAATTAACATTCAA 2157
 |||||||
 QY 1681 AATAATAAGATTTTGTATAAATGCTAAGACGCTGATGTGATATATATAGACGCTCAG 1740
 |||||||
 Db 2158 AATAATAAGATTTTGTATAAATGCTAAGACGCTGATGTGATATATATAGACGCTCAG 2217
 |||||||
 QY 1741 GCGATTAGTTTAAAGGAGCAGCTTCACGCTCTTTCTTAGCTAAGCTTAGTCTACAGATG 1800
 |||||||
 Db 2218 GCGATTAGTTTAAAGGAGCAGCTTCACGCTCTTTCTTAGCTAAGCTTAGTCTACAGATG 2277
 |||||||
 QY 1801 ATAGATTATCACTCTTAAATAATCAATTTACCGGCTCTTGTCTAAGCGTGGTGGTGAATTA 1860
 |||||||
 Db 2278 ATAGATTATCACTCTTAAATAATCAATTTACCGGCTCTTGTCTAAGCGTGGTGGTGAATTA 2337
 |||||||
 QY 1861 TTCCTTTGAGTCCAAACAACTTATGACCGTACAGGTGACGTTGGAGCTAAACAAATAG 1920
 |||||||
 Db 2338 TTCCTTTGAGTCCAAACAACTTATGACCGTACAGGTGACGTTGGAGCTAAACAAATAG 2397
 |||||||

RESULT 2
 AE008674/c
 LOCUS
 DEFINITION Rickettsia conorii Malish 7, section 106 of 114 of the complete genome.
 ACCESSION AE008674 AE006914
 KEYWORDS AE008674.1 GI:15620409
 SOURCE
 ORGANISM Rickettsia conorii.
 Rickettsia conorii.
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
 1 (sites)
 Oyata,H., Audic,S., Barbe,V., Artiguenave,F., Fournier,P.E.,
 Raoult,D. and Claverie,J.M.
 Selfish DNA in protein-coding genes of Rickettsia
 Science 290 (5490), 347-350 (2000)
 20485642
 11030655
 2 (sites)
 Oyata,H., Audic,S. and Claverie,J.-M.
 Selfish DNA and the origin of genes
 Science 291 (5502), 252-253 (2001)
 3 (bases 1 to 10078)
 Oyata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V.,
 Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and
 Raoult,D.
 Mechanisms of evolution in Rickettsia conorii and R. prowazekii
 Science 293 (5537), 2093-2098 (2001)
 21442074
 11557893
 4 (bases 1 to 10078)
 Oyata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V.,
 Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and
 Raoult,D.
 Direct Submission
 Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine,
 CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean
 Moulin, Marseille Cedex 05 13385, France
 A public version of R. conorii genome database is accessible at
 http://igs-server.cnrs-mrs.fr/. The database intends to provide
 updated data.. Annotation of the genome is an ongoing task whose
 goal is to make the genome sequence more useful. Comments to the
 authors are appreciated.

FEATURES
source

gene

CDS

Location/Qualifiers

1..10078
 /organism="Rickettsia conorii"
 /strain="Malish 7"
 /db_xref="taxon:781"
 complement(93..6158)
 /gene="rompA"
 /note="RC1273"
 complement(93..6158)
 /gene="rompA"
 /codon_start=1
 /product="190-KDa cell surface antigen"
 /protein_id="AA03811.1"
 /db_xref="GI:15620410"
 /translation="MANISPKLFOKAIOGQKAAALFTTSTAAIMLSSSGALGIASVGV
 AINNNAAFSNDVGNWNEITAVGNTTPARGPQNNWMAFTYGGDYITADVAADHIIT
 ATNVADTTPGLNIAQNTVVGSIIVTGGNLLPITITAGSLTLNGNDAADAHNGFCAPA
 DNYTGLNIALGGANAALIIOQAAPAKITLAGNIGGIIIVTKDAAIIGNTIGNAL
 ATNVGAGIATLEGAIKATTTKLTNAASVLTITNVAVLGTADNTGVDNVGNLNL
 NGALSOVTGNIGNTNATATISVGAGKATLGGAVIKATTTKLTDNASAVTFNPVVVTG
 AIDNTGNNGIVTFTGDSIVTGNIGNTNATATISVGAGKATLGGALIKATTTKLTDN
 ASVTFNPVVVTGAIDNTGNNGIVTFTGDSIVTGNIGNTNATATISVGAGKATLGG
 GAIKATTTKLTDNASAVTFNPVVVTGAIDNTGNNGIVTFTGDSIVTGNIGNTN
 LATISVGAGKATLGGAIKATTTKLTDNASAVTFNPVVVTGAIDNTGNNGIVTFT
 GDSIVTGNIGNTNATATISVGAGKATLGGAIKATTTKLTDNASAVTFNPVVVTGAI
 DNTGNNGIVTFTGDSIVTGNIGNTNATATISVGAGKATLGGAVIKATTTKLTNNAAS
 VLTITNNAVLGTADNTGVDNVGNLNGALSOVTGNIGNTNATATISVGAGKATL
 GGAVIKATTTKLTDNASAVTFNPVVVTGAIDNTGNNGIVTFTGDSIVTGNIGNTN
 ALATNVGAGLLRVQGGVKSNTITLNDNASAVTFNPVVVTGAIDNTGNNGIVTFT
 TGDSIVTGNIGNTNATATISVGAGKATLGGAIKATTTKLTDNASAVTFNPVVVTGA
 IDNTGNNGIVTFTGDSIVTGNIGNTNATATISVGAGKATLGGAVIKATTTKLTNNAAS
 TLENGPLDGGNAIPYFYGAIANGNAILNVTKLTLLTAHLIGIVPAEINIGAGNL
 FADASAGDVTILNAQDILHFRALDSALVLSLTGTVGNVNNILLAADLVAPGDEGTVF
 DGGVNLNIGSNVAGAAINTDGVGNKFTLLIYNAVITTDVNLLEGIONVLIINNAD
 FTSFAPNAGTIOINDATYITDANNGLNIPAGNIKFAHADQILIONSGNDRTITL
 GANIDPNDDEGIVILNSVTAGKLTITAGKTGGAKHLODIVFKGEGDCTAGTTFN
 TITNVIDITGOELGATANVLFKDAVOLTOGTGNIAGFDNFKNGTITLNNVNA
 GTYKNTGCTNNGTLLIVGLASNLNRVNGIAMLKVGAGNVTIAGKNVKIGELQGTGNT
 LTLPAPFKLTGSIINKTGQALKLNFNGSVGVVGTAAANSVGGDITTAGATSFASVNV
 AKGTATLGGTTSFAGTNTGAVTLAKGSIITSFANKVTATSFANSATINFGSLAFN
 SNITGSGTTLTLAGANQVYTGDTGTDTLTNTTDFGAAGSGGNILKSGSLDLSGV
 SNIAVVTATNFDMNNISPDTKYVISAETAGGLKPTPKENKVTITINNDRVDFOTFD
 ASTLTLEAEDIAAGVIDEPAGGPLANIPNAANTIKSLEMDAPNGSDARQAFNE
 GLMPTLOEADATHLAQDVYKPSDTIAAVNVQVYASNISSNITALNARMKQVQAGNK
 PVSGDEMDAKFGAWISPPVGNATQKMCNSISGYKSDTGGTIGFDGFGVDDLVGLL
 AVTRADTDIKLNNKTDGKNKNESYISLYGLYSPYENFVEAIASYSDNKRISKRS
 RVATLTLETGYQTANGKYSESYTGQLMAGTYTMMSENILTLPLAGLRYSTIKDKSY
 KETGTTYNLTIVKGNKNTFDGLLAKVSSNINVNEIVLTPELAMVDYAFKNKVSAT
 DARLQGMTAPLPTNSFKQSTSFDPGVGVGTAKHKMEYGINYDNTNIGSKYFAQOGSVK
 VRVNF"
 complement(6777..9011)
 /gene="ftsK"
 /note="RC1274"
 complement(6777..9011)
 /gene="ftsK"
 /codon_start=1
 /product="cell division protein ftsK homolog"
 /protein_id="AA03812.1"
 /db_xref="GI:15620411"
 /translation="MLYINKILSNKKVQAVILGILGIVIVLTSTYNIIDDPSPFSVT
 TEYHNLVIGSVLSDCLVQFFGLAFLIPLACFVWGRNCWGRGSRFMRFMVLL
 ALVSSSTLLSKIKLEFIPANAGGAIGIANSFFRNTNQLLYLLLIFFITLIVLVEI
 KFTLSLNFIIKLGNLIYRIQSFVHLNFSRLSSRIPTKNDKINTSSYQKPVSEK
 VKPEARSVPANPKFSPKSPVPSKISQSEIAPLPSLLDRDPEKHVKGASLSLQK
 KABELLTVLNDVFGVGOIININOCVPVYQVEFEPAGTKTSRVVGLSDDLARSLSLQ
 TRIAYIPGNVLGIELPNKOREFELKELIETPEYQDKSTLLPLVLCKDLQAGPLVAD
 LAKMPLHLLVAGTTGSGKSVGINVMVLSLLYRTPPEECRFTIMDKMLELSAYDGIPLH
 LTPVTEPSKAVVALKWKAKEMRYRMNSNIGVKNAGIYNAKLEAVKNERIIRSI
 QTFDEPTGKIYETVMKMEKLPYIVIVDEADLMLVAGKDIEMLIQRLAARAA
 GIHINATQRPVSDVITGVIKANFPSPISFKVTSKIDSRITLGBQSGEQLIGMDMLF
 MGSTSKISRVHGFVNEAEITEYLKESGTPEYISAVTEQPEEDDDSGIDIGDGTSD

gene

CDS

```

EVLYKAVOIVDERKSSYSYIQRSLRIGYKNKAANLVEKMEKEGIVPPNHTGKREIL
LPER"
gene      complement(9645..9968)
CDS       complement(9645..9968)
          /gene="RC1275"
          /codon_start=1
          /transl_table=11
          /product="unknown"
          /protein_id="AAL03813.1"
          /db_xref="GI:15620412"
          /translation="MGVRLRLPALSVAAHSHPLKFIQITTIIVQSFPMISGYMA
MIOSNYTSIKIKFWKRYRLPYIFCILVTFLOOKFSLNSVIFHPRLVFLIFAN
LTMLL"
BASE COUNT 3031 a 2134 c 1617 g 3296 t
ORIGIN

Query Match      32.1%; Score 615.4; DB 1; Length 10078;
Best Local Similarity 65.7%; Pred. No. 1.3e-103;
Matches 958; Conservative 0; Mismatches 471; Indels 30; Gaps 3;

QY 1 ATGCGGAATATTTCTCTAAATATTTTCAAAAAGCAATTCAAAAGCTCTTAAAGCTGCT 60
DB 1 ATGCGGAATATTTCTCTAAATATTTTCAAAAAGCAATTCAAAAGCTCTTAAAGCTGCT 6099
QY 61 TTATTCAACACCTCAACCGCAGCGATAAAGCTTAACCGGTAGTGGAGTCTTGGTGTGCA 120
DB 61 TTATTCAACACCTCAACCGCAGCGATAAAGCTTAACCGGTAGTGGAGTCTTGGTGTGCA 120
QY 6098 TTATTCAACACCTCAACCGCAGCGATAAAGCTTAACCGGTAGTGGAGTCTTGGTGTGCT 6039
DB 6098 TTATTCAACACCTCAACCGCAGCGATAAAGCTTAACCGGTAGTGGAGTCTTGGTGTGCT 6039
QY 121 AGAACGCTACTGCTGATGTGTCAGAGCTTGCAGCGGACAAATATAGT-----171
DB 121 AGAACGCTACTGCTGATGTGTCAGAGCTTGCAGCGGACAAATATAGT-----171
QY 6038 GTTTCAGGTGTTATGCTACTAATAATATGTCAGCATTTAGTATGTTGGCAATAAT 5799
DB 6038 GTTTCAGGTGTTATGCTACTAATAATATGTCAGCATTTAGTATGTTGGCAATAAT 5799
QY 172 -----CCTGAGCGCGTGTGTTTGTAGCGGTTCTACTTTTACAAATATACCGGT 219
DB 172 -----CCTGAGCGCGTGTGTTTGTAGCGGTTCTACTTTTACAAATATACCGGT 219
QY 5978 TGGAAATGAGATAACGGCTGACGGGTAGCTAATGTTACTCTGCTCGCGTCTCAAAC 5919
DB 5978 TGGAAATGAGATAACGGCTGACGGGTAGCTAATGTTACTCTGCTCGCGTCTCAAAC 5919
QY 220 GCTTTTACGGTACTGATGCTGACGTAAGTGTGTCGATAGTATTAATAATTTTGA 279
DB 220 GCTTTTACGGTACTGATGCTGACGTAAGTGTGTCGATAGTATTAATAATTTTGA 279
QY 5918 AATTGGGCATTACTTACGGTGTGATTATATCTATCTGATGATGATGATGATGATGAT 5859
DB 5918 AATTGGGCATTACTTACGGTGTGATTATATCTATCTGATGATGATGATGATGATGAT 5859
QY 280 GCAGGTCTTTTTCAGTAACCTGCTGATATTTTATTTAGTTTCACTGCTGATGATGATGAT 339
DB 280 GCAGGTCTTTTTCAGTAACCTGCTGATATTTTATTTAGTTTCACTGCTGATGATGATGAT 339
QY 5858 ATTACGGCTAATAATGTTGCGGATACCTACCTCCATAGGTCTAATATGCTCAAAATACC 5799
DB 5858 ATTACGGCTAATAATGTTGCGGATACCTACCTCCATAGGTCTAATATGCTCAAAATACC 5799
QY 340 GCTAATAAATCTGCAGTATATTTATGATGTTTAACTTAACTTAACTTAACTTAACTTAACT 399
DB 340 GCTAATAAATCTGCAGTATATTTATGATGTTTAACTTAACTTAACTTAACTTAACTTAACT 399
QY 5798 GTCGTTGGTTCGATTAAACGGGAGGTAACTTGTGCTGTTACTTATCTGCGGCAAA 5739
DB 5798 GTCGTTGGTTCGATTAAACGGGAGGTAACTTGTGCTGTTACTTATCTGCGGCAAA 5739
QY 400 ACTGACGCTACGGTGCAGAAATCCTCGGTTGTTATTCAGGTGGACAAAGCTGCTGTAAT 459
DB 400 ACTGACGCTACGGTGCAGAAATCCTCGGTTGTTATTCAGGTGGACAAAGCTGCTGTAAT 459
QY 5738 AGCTTAACCTTAAACGGTAAATATGCTGATGCTGCAAAATCATGTTGTTGCTCCTGCC 5679
DB 5738 AGCTTAACCTTAAACGGTAAATATGCTGATGCTGCAAAATCATGTTGTTGCTCCTGCC 5679
QY 460 AATACATATACCTGCTTAGTATATATACCTAGTGGAGCGAATGCCGTTTGCATATT 519
DB 460 AATACATATACCTGCTTAGTATATATACCTAGTGGAGCGAATGCCGTTTGCATATT 519
QY 5678 GATAATATACAGGTTAGGAATATAGCTTTAGGGGAGCGAATGCTGCACTAATAT- 5620
DB 5678 GATAATATACAGGTTAGGAATATAGCTTTAGGGGAGCGAATGCTGCACTAATAT- 5620
QY 520 GCTTTCAGATCCAGATGATTAGGACCAATAACGCTTGCAGGAATATAGATGAGGAGGT 579
DB 520 GCTTTCAGATCCAGATGATTAGGACCAATAACGCTTGCAGGAATATAGATGAGGAGGT 579
QY 5619 -----ACAATCTGACGCCCGGCAAGATATACACTTGCAGGCAATATATGAGGAGGT 5565
DB 5619 -----ACAATCTGACGCCCGGCAAGATATACACTTGCAGGCAATATATGAGGAGGT 5565
QY 580 ATAATACTGACATACAGATGCTGCAATTAACCGAACAATAGTATATACTAATCCGGCA 639
DB 580 ATAATACTGACATACAGATGCTGCAATTAACCGAACAATAGTATATACTAATCCGGCA 639
QY 5564 ATAATACTGCAAGACAGATGCTGCAATTAACCGAACAATAGTATATACTAATCCGGT 5505
DB 5564 ATAATACTGCAAGACAGATGCTGCAATTAACCGAACAATAGTATATACTAATCCGGT 5505
QY 640 GCTCAATAGCATTTGGACAGTACGCTTCTTCTTGGAGGCGAGTTATTAAGCCACT 699
DB 640 GCTCAATAGCATTTGGACAGTACGCTTCTTCTTGGAGGCGAGTTATTAAGCCACT 699
QY 5504 GCAACAGTGAATGAGGACAGGTATAGCACGTTTAGAGGAGCGGATTTAATGAGTACT 5445
DB 5504 GCAACAGTGAATGAGGACAGGTATAGCACGTTTAGAGGAGCGGATTTAATGAGTACT 5445
QY 700 ACGACTAAGTAAACAAATGAGCGCGGATTTTAACTTAACTTAACTTAACTTAACTTAACT 759
DB 700 ACGACTAAGTAAACAAATGAGCGCGGATTTTAACTTAACTTAACTTAACTTAACTTAACT 759
QY 5444 ACGACTAAGTAACTGCTGCTGGTATTAACCTTAACTTAACTTAACTTAACTTAACTTAACT 5385
DB 5444 ACGACTAAGTAACTGCTGCTGGTATTAACCTTAACTTAACTTAACTTAACTTAACTTAACT 5385

```

```

QY 760 ACAGGTGCTGTTGATTAACACACACAGCGGTGATGATGATGCTGCTTAAATTTAAACGGA 819
DB 5384 ACAGGTGCTGTTGATTAACACACACAGCGGTGATGATGATGCTGCTTAAATTTAAATGTT 5325
QY 820 GCGTTGAGTCAAGTAACTGGAATATAGTAAATACAAATTCATAGGCAATAGTGTGA 879
DB 5324 GCATTAAAGTCAAGTAACTGGAATATAGTAAATACAAATTCATAGGCAATAGTGTGA 5265
QY 880 GGAGCAGGTACGCTACGTTAGGGGAGCGGTTATTAAAGCCACTACGACTAAGTTAAACG 939
DB 5264 GGAGCAGGTAAACCCAGCTTAGGGGAGCGGTTATTAAAGCTACTAGGCAATAGTGTGA 5205
QY 940 AATGCAAGCGTGGTATTAACACTTACAAATGCAAGTATTAACAGGTGCTGTTGATAACAC 999
DB 5204 GATAATGCGTCAGCAGTGACATTTACGAATCCTGTAGTGGTGACCGGAGCATAGATAAT 5145
QY 1000 ACAGCGCGGTGATTAATGATGCTGCTGTAATCTAAGTGGAGCATTAAGTCAAGTAAACGGG 1059
DB 5144 ACCGGTAATGCAATTAATGCTATAGTAACGTTTACCGGTG---ATAGTACAGTAACGTGGC 5088
QY 1060 AATATAGGTAAATACAAATTCATTAGCCACGATAAATATAGGACGAGTGTGGCTACCTTG 1119
DB 5087 AATATAGGTAAATACAAATTCATTAGCCACGATAAATATAGGACGAGTGTGGCTACCTTG 5028
QY 1120 GATGGACGCGTTTAAAGCTACTAGCAGTAAAGTAAACAGATGATGCTGAGTATTAATA 1179
DB 5027 GGGGAGCGGATTATTAAGCTACTAGCAGTAAAGTAAACAGATGATGCTGAGTATTAATA 4968
QY 1180 TTTACGAATCCTGCTAGTAAACCGGAGCAATAGATAATACCGGTAAATGCAATAAAGGT 1239
DB 4967 TTTACGAATCCTGCTAGTAAACCGGAGCAATAGATAATACCGGTAAATGCAATAAAGGT 4908
QY 1240 GTGCTAATCTTTACCGGAGCAAGTACGGTAACCGATTAATATAGGTAAACCGGAGTATTA 1299
DB 4907 ATAGTAACTGTTACCGGAGCAAGTACGGTAACCGATTAATATAGGTAAACCGGAGTATTA 4848
QY 1300 GCAGAGTAACTGAGGAGCAGGTTTCTGCAATTAACAGCGGAGTAGTAAAGCGAAT 1359
DB 4847 GCACGATAAGTGTAGGAGCAGGTTTAAAGCCAGCTTATAGGCGGAGCGATTTATTAAGTACT 4788
QY 1360 GCAATAAATTTACGGAATATGCTAGTAAACATTTTACCGGTGATAGTACGGAATCA 1419
DB 4787 ACAGCTAAGTTACGGAATATGCTAGTAAACATTTTACCGGTGATAGTACGGAATCA 4728
QY 1420 GGTAGTAACTGCTGATACAG 1438
DB 4727 GGAGCGATAGATTAATACCG 4709

```

```

RESULT 3
RIRANT190K
LOCUS      R.rickettsii 190 kd antigen gene, complete cds.
DEFINITION M31227.1 GI:152465
ACCESSION M31227.1
VERSION    cell surface antigen.
KEYWORDS   R.rickettsii (strain R) DNA, clones p1100 and pGAM21.
SOURCE     Rickettsia rickettsii
ORGANISM   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
            Rickettsiaceae; Rickettsiinae; Rickettsia; spotted fever group.
            1 (bases 1 to 7088)
REFERENCE   Anderson,B., McDonald,G., Jones,D. and Hagnery,R.
            Protective protein antigen of Rickettsia rickettsii has tandemly
            repeated: near-identical sequences
            Unpublished (1990)
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by B.Anderson, 12-JAN-1990.
FEATURES
            source
            1..7088
            /organism="Rickettsia rickettsii"
            /db_xref="taxon:783"
            -35_signal
            3..8
            -10_signal
            26..31

```



```

RBS      58..62
          /note="ribosome binding site"
CDS       70..6819
          /note="cell surface antigen"
          /codon_start=1
          /transl_table=11
          /protein_id="AAA26380.1"
          /db_xref="gi:152466"
          /translation="MANISPKLFKALIOGGLKAALFTTSTAAILMSSSGALGVATGVI
          ATNNNAESNVGNWNNNEITAAAGVANGTTPAGGPONNNAFTYGGDYTTADAADRIIK
          AINVAGTTPVLNITNTQNTVVGSIITKGNLPLVTNAGKSLTLNGNNAVAANHGFDAPA
          DNYTGLGAGTALGGANAALIIQSAAPSKITLAGNIDGGIITVKTDAALNGTIGNTAL
          ATVNGAGTATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLN
          NGALSOVTDIGTNSLATISVGAGTATLGGAVIKATTTKLTDAASAVKFTNPVVTC
          ALDNTGNANGIVFTGNTSVGNVNTNALTATVNGAGLLOVQGVVYKANTINLTDN
          ASAVTTPNPVVTCGAIDNTGNANGIVFTGNTSVGNVNTNALTATVNGAGLLOVQGVVYKANTINLTDN
          GAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLNNGALSOVTDIGTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLN
          TFGNTSVTDIGTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLN
          ITNAVSAVKFTNPVVTCGAIDNTGNANGIVFTGNTSVGNVNTNALTATVNGAGLLOVQGVVYKANTINLTDN
          ATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLNNGALSOVTDIGTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLN
          DIGNTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLNNGALSOVTDIGTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLN
          DNYGVLNNGALSOVTDIGTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLNNGALSOVTDIGTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLN
          TNANAVLTGAIDNTTGDGNGVNLNNGALSOVTDIGTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLNNGALSOVTDIGTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLN
          IKATTTKLDAASAVKFTNPVVTCGAIDNTGNANGIVFTGNTSVGNVNTNALTATVNGAGLLOVQGVVYKANTINLTDN
          VNGAGLLOVQGVVYKANTINLTDN
          TVTGNVNTNALTATVNGAGLLOVQGVVYKANTINLTDN
          GNANGIVFTGNTSVTDIGTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLNNGALSOVTDIGTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTNANAVLTGAIDNTTGDGNGVNLN
          NGPLDGGKALIPYFGAIANGNNAILNNTKLTASHLTIGTVAEINIDGAGNLTID
          ASVGTITLNAQNTINFRADSVLNSLTGVTNLLIADLVAPGADGTVWNGV
          NGLNVSNAQNTINFRADSVLNSLTGVTNLLIADLVAPGADGTVWNGV
          TAENAGAIQINDATYIDANNGLNIPAGNIOFADHQAQVLVNLNSGNDRTITLIGANI
          DPNDDEGIIVLNSVAGKLTITAGKTFGGAKHQLTTLFKGAGCSTAGTITFNINI
          VLDITDQLELATANVLFNDVOLTQTNIGLDFNARKNGMVTLNANNVAVAGVQ
          NTGNTNGLTILVGLASNLNVNGTAMLVKAGNVTIAKGGVKIGIEIGTNTLILP
          AFHNLGTSINTGQALKNFMNGSVSVGVTAAANSVDITTAGATSFASVNAKGT
          ATLGSTTSFANTVGTAVTLAKGITSFAKNVTATSEVANSATINFSNLSAFNSNIT
          GGGTTLTGANOVTVTGTSFTDLTLNTPDGAAGSGNLIKSGSLDLSGVSTLA
          LVVATNFDMNISPDKTVIISAFTAGLPTSKENVKIINDNREVDVTFDASTL
          TLFEADIAADIDGDFAPGGLANIPNAANIKKSLMEDAPNGDARQAFNNGLMT
          PLOEADATHLIQDVKPSDTIAVNNQVVAASNSNITLALNMRDMKQVOSGNKPVSS
          GDEMDFAGNWKISPFQNGATQKMCNSISYKSDPTGGTIGDFGVSDDLALGLAYTR
          ADTDLKLNKNTGKQVNESNIYSLYGLNYPENLFVEATASYSDNRIRSKSRVIA
          TLTETVQNGKNTYKQKSEYTGOLMAGTYMMPENILNLTPLAGLRYSTIKQKQKETG
          TTYQNLTVKGNKNTYKQKSEYTGOLMAGTYMMPENILNLTPLAGLRYSTIKQKQKETG
          OGMRAPLTPNSFKQSKTSFDFGVGVGTAKHKKMEXRINDTNTNIGSKYFAQQGSVKVRVN
          P"
          703..3564
          /note="432 bp repeats"
          6839..6876
          /note="transcription termination signal"
BASE COUNT 2380 a 1125 c 1618 g 1965 t
ORIGIN
repeat_region
misc_signal
Query Match
Best Local Similarity 31.1%; Score 597.6; DB 1; Length 7088;
Matches 982; Conservative 0; Mismatches 518; Indels 36; Gaps 4;
QY 1 ATGGCGCAATATTTCTTAAATATTTTCAAAAGCAATTTCAAAAGGCTTTAAACTGCT 60
DB 70 ATGGCGCAATATTTCTTAAATATTTTCAAAAGCAATTTCAAAAGGCTTTAAAGCGCT 129
QY 61 TTATTCACCACCTCAACCCGACGATATGCTTAAACCGGTAGTGGAGTCCCTGCTGCTGA 120
DB 130 TTATTCACCACCTCAACCCGACGATATGCTTAAACCGGTAGTGGAGTCCCTGCTGCTG 189
QY 121 AGAACCGTAACTGCTGATGGTCAGAGCTTCGACCGCGGAACAATATAGTCTCTGAGCC 180
DB 190 ACAGGTGTTTATGCTACTATAATAATGACGATTTAGTAAATATCTTGGCAATAAT 249
QY 181 GTGCTTTTGTAGCGGTTCTACTTACAAATATACCGGTGCTTTTACGGTTTACTGATGT 240

```

```

DB 250 TGAATGAGATAACGCGCTCAGGGGTAGCTAATGTTACTCTCTCGCGGCTCTCAAAAC 309
QY 241 GACGTAAGTGTTCGTGCATATAGATTAAATAATTTTGCAGCAGGTCTTTTTTCAGTAAT 300
DB 310 AATTTGGGCAATTTACTTTACGGTGTGATTATCTGTCTCAGTACGAGCCGATCTGATT 369
QY 301 GGTGATATTTCTATAGTTTACGTTAGTAGTACGGGAGGAGCAATAAATTCAGTAAAT 360
DB 370 ATTAAGGCTATAAATGTTCCGGGTACTACTCCGCTAGGTCTAAATATTTACTCAAAAT 429
QY 361 ATTGATG-----ATGTTTAACTTAACTTTAAACAGGTACCGGT 399
DB 430 GTCGTTGGTTCGATTATAACGAAAGTAACTTCTGCTGTTACTCTTAATGCCGCAAA 489
QY 400 ACTGACGCTCAGGTGCAAAATCCTGCTGTTTATTTCCAAAGTGGGCAACGCTGCTGCTAAT 459
DB 490 AGCTTAACTTTAAATGTTAATGCTGTGCTGCAAAATCATGTTTGTATGCGCCTGCC 549
QY 460 AATACATATCTGCTTTAGGTAATACTCTAGGTGGAGGCAATCCCGTGTGACTATT 519
DB 550 GATAATTATACAGTTTAGGAATAATAGCTTTAGGGGGAGCGAATGCTGCATAATAT- 608
QY 520 GCTTCAGATCCAGATGTTATTAGGACCAATAACGCTTTGCAGGAAATATAGATGGAGGAGT 579
DB 609 -----ACAATCTCGAGCTCCGTCARAGATAACACTTGCAGGAAATATAGATGGAGGAGT 663
QY 580 ATAATAACTGACAAATACAGATGCTGCCATTACGGGAACAATAGGTAAATCTAATCCGGCA 639
DB 664 ATAATAACTGTCAAGACAGATGCTGCCATTACGGGAACAATAGGTAAATCAAAATGCAATTA 723
QY 640 GCTCAAAATAGCATTTGAGCAAGTACGCTTTCTCTTGGAGGGGCGACTTTTAAAGCCACT 699
DB 724 GCAACAGTGAATGTAGGACACAGTACAGCCAGCTTAGGGGGAGCGGTTTATTAAGTACT 783
QY 700 ACAGCTAAGTTTAAACAAATCAGCGCGGTATTAACTTACAAATGCAAAATGCAATTAAGTATTA 759
DB 784 ACAGCTAATTAACGAATGCTGCTCGGTATTAACTTACAAATGCAAAATGCAATTAAGTATTA 843
QY 760 ACAGTGTCTGTGATAACACACGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 819
DB 844 ACAGGTGCGGATGATAACACACGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 903
QY 820 GCGTTGAGTCAAGTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 879
DB 904 GCATTAAGTCAAGTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 963
QY 880 GGAGCAGGTACGCGTACGTTAGGGGAGCGGTTTATTAAGCCACTACGACTAAGTTAAACG 939
DB 964 GGAGCAGGTACGCGTACGTTAGGGGAGCGGTTTATTAAGCCACTACGACTAAGTTTGA 1023
QY 940 AATGACGCTGCGTATTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 999
DB 1024 GATGCTGCTGACGAGTGAATTTACGAATCTGCTAGTGGTACTGGAGCGGATAGATAAT 1083
QY 1000 ACAGCGGTGATAAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1059
DB 1084 ACCGGTAAATGCAAAATATGCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1060 AATATAGTAAATACAAATTCATTTAGCCAGATAAATAATAGGACGAGGTGGTGGTACTGCTG 1119
DB 1141 AATGAGTAAATACAAATTCATTTAGCCAGATAAATAATAGGACGAGGTGGTGGTACTGCTGCTG 1200
QY 1120 GATGAGCGGTTTATTAAGCTACTACGACTAAGTTTAACTTAACTTAACTTAACTTAACTTAACTTAA 1179
DB 1201 CAAAGTGGAGTGGTAAACCAATCAATTAACCTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1260
QY 1180 TTTTACGAATCTGCTAGTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1239
DB 1261 TTTTACGAATCTGCTAGTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1320
QY 1240 GTGGTAACTTTTACCGGAGCAAGTACGGTAAACGATAAATAATAGGTAAACGAGCGGATATTA 1299
DB 1321 ATAGTAACTTTTACCGGATATAGGTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 1380

```


Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.
 1 (bases 1 to 6361)
 Croquet-Valdes, P.A., Weiss, K. and Walker, D.H.
 Sequence analysis of the 190-kDa antigen-encoding gene of
 Rickettsia conorii (Malish 7 strain)
 Gene.140 (1), 115-119 (1994)
 94171067
 8125327
 2 (bases 1 to 6361)
 Croquet-Valdes, P.A.
 Direct Submission
 Submitted (23-AUG-1993) P.A. Croquet-Valdes, University of Texas
 Medical Branch at Galveston, Pathology, 11th Street and Texas
 Avenue, Galveston, Texas 77555, USA
 Location/Qualifiers
 1..6361
 /organism="Rickettsia conorii"
 /strain="Malish 7"
 /db_xref="taxon:781"
 3..8
 26..31
 59..63
 72..6137
 /codon_start=1
 /transl_table=11
 /product="190-kDa antigen (rOmpA)"
 /protein_id="AA17405.1"
 /db_xref="GI:467814"
 /translation="MANISPKLFKKAIOQGLKALFTSTTAAILMSSGALGIASVGV
 IATNNAFSDVNGNNEITPAAGVNGTPARQPNNAFTYGGDYTVADVAHHIT
 ANVDATPIGLNAQNTVGVIIITGNLPLVITNAGKSLITLNGNDAANHHGADPA
 DNYTGLNIALGANAALIIOQAPAKITLAGNIGGGLITVKTDAALNGTIGNAL
 ATVNGAGIATLEGAIKATTKLTNAAVLITNVRNLFGADINTGYDYNVLNL
 NGALSOYTGNTNATATISVGAKATLGAVIKATTKLTNDASAVTFNPVVVVG
 AIDVTGNNGIVTIGDSTVGNIGNATLVNAGLLQVGVVYKANTINLND
 ASAVTFNPVVVVGADIDNTGNNGIVTIGDSTVGNIGNATLVNAGLLQVGVVYKANTINLND
 GAIITKATTKLTNDASAVTFNPVVVVGADIDNTGNNGIVTIGDSTVGNIGNATLVNAGLLQVGVVYKANTINLND
 LATISVGAKATLGAIKATTKLTNDASAVTFNPVVVVGADIDNTGNNGIVTIGDSTVGNIGNATLVNAGLLQVGVVYKANTINLND
 GDSVTGNTGNATATISVGAKATLGAIKATTKLTNDASAVTFNPVVVVGADIDNTGNNGIVTIGDSTVGNIGNATLVNAGLLQVGVVYKANTINLND
 DNTGNNGIVTIGDSTVGNIGNATLVNAGLLQVGVVYKANTINLND
 VLTITNVAVLGADINTGYDYNVLNAGLLQVGVVYKANTINLND
 GGAVIKATTKLTNDASAVTFNPVVVVGADIDNTGNNGIVTIGDSTVGNIGNATLVNAGLLQVGVVYKANTINLND
 ALATISVGAKATLGAIKATTKLTNDASAVTFNPVVVVGADIDNTGNNGIVTIGDSTVGNIGNATLVNAGLLQVGVVYKANTINLND
 TGDSTVGNIGNATLVNAGLLQVGVVYKANTINLND
 IDNTGNNGIVTIGDSTVGNIGNATLVNAGLLQVGVVYKANTINLND
 IDNTGNNGIVTIGDSTVGNIGNATLVNAGLLQVGVVYKANTINLND
 TLEFNGPLDGGKAIPIYKGAJANGNALINLVNFKSLTASHLITIGVFAEINIGAGNL
 FAIDASAGDVTILNAQDIHFALDSALVSLNLTGVNVLNLAADLVAPGVDEGTVPF
 DGVNGLNIGSVNAGARNIGDVGKNTFLLIYNAVITDDVNLGIONVLLNNAD
 FTSSTFAGNTIQTINDATYITDANGNLINPAGNIQFAHADAQILONSSGNDITFL
 GANTDPNDDEGIVILNSVTAGKRLTAGKTGGAGHKLQDIVEFGEDDFGTAGTTFN
 TTNITVDITGOLELGATNVNLFKDAVQLTQTNIGGIDFNKAGNTVTLNNVNA
 GTVNTGTNGTLTILVLGASNLNRVNGIAMLKVGAGNVITAKGNGVKIIGELGTGNT
 LTLPAHFELGSIKNTGGQALKLNFNGGSGVGVVGAANSVGDITTAGATSPASSVN
 AKGTALGGTTSFANTFNTGAVTLAKGSIISFAKNVTATSFVANSATINFGNSLAFN
 SNITGSGTTLFGANQVYTGSTFDITLNTFDGAAGKSGNLLKSGTDLGSLV
 SNLALVTATFNDDNNISPDTKYVISAETAGLKPPTKENVKITINNDNFVDFPD
 ASTLTLEADIAADIDGDFAPGPLANIPIAANKSLEMDAPNGDARQAFNF
 PLMSDLEADATILMODVYKPSDTIAAVNNOVVASINSNITATNARMKQVQNGK
 PVSSGDEMDAKFANISPVFNATQKNSISYKSDTGGTIGFGFVSDDLVGLR
 AYTRADIDILKNNKTDGKNKVESNITSLGLYSPYENLFEVIAISYDNKIKRSR
 RVITLLETGVYQFANGKYSESYGQLMVGYPYMLPENLPLAGLRYSTIKDKSY
 KETGTYTQKNTNFTDGLGKSVSSNINVAIPLTPELYAMDYAFKKNVSAT
 DARLQGITPLSTNSFKQSKTSFDVGVGVTAHKMKMEYRINYDINIGSKYFAQOGSVK
 VRVNF"

repeat_region 705..2882
 /rpt_type=tandem
 terminator 6156..6194
 BASE COUNT 2143 a 1015 c 1431 g 1772 t
 ORIGIN

Query Match 30.68; Score 587.8; DB 1; Length 6361;
 Best Local Similarity 64.94; Pred. No. 1.7e-98;

	Matches	947;	Conservative	0;	Mismatches	482;	Indels	30;	Gaps	4;
QY	1	ATGCGCAATATTTCTCTAAATATTTTCAAAAGCAATTTCAAAAGGCTTTAAACTGCT	60							
Db	72	ATGCGCAATATTTCTCTAAATATTTTCAAAAGCAATTTCAAAAGGCTTTAAAGCGCT	131							
QY	61	TTATTCCACCCTCAACCGCAGGATATGCTAAACCGGTAGTGGAGTCTTGGTGTGCA	120							
Db	132	TTATTCCACCCTCAACCGCAGGATATGCTAAACCGGTAGTGGAGTCTTGGTGTGCA	191							
QY	121	AGAACCGTAACCTGCTGAGGTGTCAGAGCTTGCAGCGGCAAAATATAGTCTCTGGAGC	180							
Db	192	GTTCAGGTGTTTACTACTATTAATATGAGCATTTAGTGATAATGTTGGCATAT	251							
QY	181	GGTGTGTTTGTAGCGGTCTTACTTACAAATACCGGTGCTTTACGGTTACTGA----	236							
Db	252	TGGAATGAGATAACGGCTGCAGGGGTAGCTAATGTTACTCTCTCGGGGTCTCAAAAC	311							
QY	237	-----TGCTGACGTAAGTGTCTGTCATAGATTTTAAATAATTTTTCACGAGCT	285							
Db	312	AATTCGGCATTTACTTACGGTGTGATTAATGCTGCTGACAGCTAGCCCATATAT	371							
QY	286	CTTTTTCAGTAACCTGTTGATATTTCAATAGGTTCAGTGTAGATACGGAGGAGCTAAT	345							
Db	372	ATTACGGCTATAAATGTTGCGGATACTACTCCCATAGGTCTAAATATGCTCAAAATACC	431							
QY	346	AAACTTGCAGTTAATATGATGATGCT-----TTAACCTTAACTTTACAGAGTACCGGT	399							
Db	432	GTCTGTTGTTGATTAATACGGGAGGTAACTTGTTCCTGTTTACTTATTAATCCGGCAAA	491							
QY	400	ACTGCGAGCTACCGTGCAGTAACCTCTCGTGTGTTTATTCAGAGTGGCAAGCTCTGCTAAT	459							
Db	492	ACCTTAACTTTAAACGGTAATGCTGATGCTGCAATCATGTTTGTATGCTCTCTGCC	551							
QY	460	AATACATATCTGCTTTAGGTAAATAACTCTAGTGTGGAGCAATCCCGGTTTACTAT	519							
Db	552	GATAATTTACAGTTTAGGAAATATAGCTTTAGGGGAGCAATGCTGCACATAATAT	610							
QY	520	GCTTCAGATCCAGATGATTAGGACCAATACCGTTCAGGAAATATAGATGAGGAGGT	579							
Db	611	-----ACATCTGACGCCCCGCAAGATAACATTTGAGGCAATATAAATGGAGAGGT	665							
QY	580	ATAAATCTGACATACAGATCTGCCATTTACGGAAACAAATAGTAATACTACTCCGCA	639							
Db	666	ATAAATCTGACAGACAGATCTGCCATTTACGGAAACAAATAGTAATACTACTCCGCA	725							
QY	640	GCTCAAAATAGCATTTGGAGCAAGTACGCTTCTCTTTGAGGGGAGTATTAAGCCACT	699							
Db	726	GCAACAGTGAATGTAGGAGCAGGTATAGCCAGCTTAGAGGGGCGGATTTAAGCTACT	785							
QY	700	ACGACTAAGTTAAACAAATGTCAGCGCGGTATTAAACACTTACAAATGCAATCAGTATTA	759							
Db	786	ACGACTAATTTAACCAATGTCGCTCGGTATTAAACCTTACAAATGCAATCAGTATTA	845							
QY	760	ACAGTGTCTGTTTAAACACACACAGCGGTGATGATGAGTGTCTTAAATTTAAACCGA	819							
Db	846	ACAGTGTCTGTTTAAACACACACAGCGGTGATGATGAGTGTCTTAAATTTAAACCGA	905							
QY	820	CGCTTCAGTCAAGTAACTGGAAATATAGTAATACAAATTCATTACGCAATATAGTGA	879							
Db	906	GCATTAAGTCAAGTAACTGGAAATATAGTAATACAAATTCATTACGCAATATAGTGA	965							
QY	880	GGAGCAGTACGGCTACGTTAGGGGAGCGGTATTAAAGCCACTACGACTAAGTAAAG	939							
Db	966	GGAGCAGTAAAGCCAGCGGTATTAGGGGAGCGGTATTAAAGCTACTACGACTAAGTAAAG	1025							
QY	940	AATTCAGCTCGGTATTAAACACTTACAAATGCAATGAGTGTCTTGTATTAACACC	999							
Db	1026	GATAATGCTCAGCAGTGCATTTACGATCTCTGAGTGGTACCGGAGGATGATAT	1085							
QY	1000	ACAGCGGTGATATGATGAGTGTCTGATAATCTAAGTGGAGCATGAGTCAAGTACCGGG	1059							
Db	1086	ACCGGTAATGCAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1142							

```

QY 1060 AATATAGGTATACAAATTCATTTAGCCACGATAAATATAGAGCAGGTGTGGTACCTTG 1119
D 1143 AATATAGGTATACAAATTCATTTAGCCACGATGAATAGAGCAGGTGTGGTACCAAGTA 1202
QY 1120 GATGCGCGGTATTAAAGCTACTACACACTAAGTTACACAGATGATCGTCACTAGTTAATA 1179
D 1203 CAAGGTGGAGTGGTAAAGCAAAATACAAATAAACTTAAACGGATAATGGTCCAGCAGTGACA 1262
QY 1180 TTTACGAATCCTGTAGTAGTAACCGGAGCAATAGATAATACCGGTAATGCCAATAAAGGT 1239
D 1263 TTTACGAATCCTGTAGTGGTACCGGAGCAGATAGATAATACCGGTAATGCCAATAAAGGT 1322
QY 1240 GTGGTAATCTTTACCGGAGCAAGTACCGTAAACCGATATATAGGTATACACGGCAGTAGTTA 1299
D 1323 ATAGTAACGTTTACCGGTGTAGTACAGTAACCTGCGCAATATAGGTATACAAATGGATTA 1382
QY 1300 GCAGAGTAAGCTAGGAGCAGGTTCTGCGCAATATACAGCGGAGTAGTAAAGCGAAT 1359
D 1383 GCCACGATAAGCTAGGAGCAGGTAAGCAAGCCACGCTAGGGGAGCGATTTATTAAGCTACT 1442
QY 1360 GCAATAAATTAACGGATAATGCGTACGTAGTAACTTTACCGGTGATAGTACGCTAACA 1419
D 1443 ACGACTAAGTTAAGGATATGCGTACGAGTGCATTTACGATCTCTAGTGGTGACC 1502
QY 1420 GGTAGTATAGGTGTACAG 1438
D 1503 GGAGCGATAGATATACCG 1521

RESULT 6
AF210694 AF210694 575 bp DNA linear BCT 11-JAN-2001
LOCUS Rickettsia sp. California 2 outer membrane protein OmpA (ompA)
DEFINITION gene, partial cds.
ACCESSION AF210694
VERSION AF210694.1 GI:12082488
KEYWORDS Rickettsia sp. California 2.
SOURCE Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
ORGANISM Rickettsiaceae; Rickettsia.
REFERENCE 1 (bases 1 to 575)
AUTHORS Raoult,D.
TITLE A new SFG rickettsia isolated from fleas
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 575)
AUTHORS Roux,V. and Raoult,D.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1999) Unite des rickettsies, Faculte de Medecine,
27 Boulevard Jean Moulin, Marseille 13385, France
FEATURES
source
1. .575
/organism="Rickettsia sp. California 2"
/strain="California 2"
/db_xref="taxon:147259"
<1..>575
/gene="ompA"
<1..>575
/gene="ompA"
/codon_start=1
/transl_table=11
/product="outer membrane protein OmpA"
/protein_id="AAG48555.1"
/db_xref="GI:12082489"
/translation="LFQKAIQKGLTKALFTTSTAATMLTSGVGLGAARTVTADGAELA
AGNIGPGAGFVAGSTLOYTGAFTVDADVSVRALDNNFAGLFSTGDISLGSVV
DTGANKLAVNIDGDLTLTLTGTAAGNAPALLFQGGGAANNRYTALGNITLGA
NAGLTASDPVLGPIITLAGNIDGGIITNT"
BASE COUNT 164 a 104 c 134 g 173 t
ORIGIN

Query Match 29.9%; Score 575; DB 1; Length 575;

```

```

Best Local Similarity 100.0%; Pred. No. 7.le-96;
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 22 TTATTTCAAAAAGCAATTCAAAAAGGTCTTAAAACTGCTTTATTCACCACTCAACCGCA 81
D 1 TTATTTCAAAAAGCAATTCAAAAAGGTCTTAAAACTGCTTTATTCACCACTCAACCGCA 60
QY 82 GCGATAATGCTAACCGGTAGTGGAGTCTTGGTCTGCTCAAGAACCGTAACCTGCTGATGGT 141
D 61 GCGATAATGCTAACCGGTAGTGGAGTCTTGGTCTGCTCAAGAACCGTAACCTGCTGATGGT 120
QY 142 GCAGAGCTTGCAGCGGAGCAAAATATAGTCTCTGGAGCCGGTGTCTTTAGCCGGGTCT 201
D 121 GCAGAGCTTGCAGCGGAGCAAAATATAGTCTCTGGAGCCGGTGTCTTTAGCCGGGTCT 180
QY 202 ACTTTACAATATACCGGTGCTTTTACGGTTACTGATGCTGACGTAAGTGTTCGTGCATTA 261
D 181 ACTTTACAATATACCGGTGCTTTTACGGTTACTGATGCTGACGTAAGTGTTCGTGCATTA 240
QY 262 GATTAAATAATTTTGCAGCAGGTCTTTTTCAGTAACTGGTGATATTTTCAATTAGGTTC 321
D 241 GATTAAATAATTTTGCAGCAGGTCTTTTTCAGTAACTGGTGATATTTTCAATTAGGTTC 300
QY 322 GTGCTAGTACGGGAGGAGCTAATAAACTTGCAGTAACTGGTGATATTTTCAATTAGGTTC 381
D 301 GTGCTAGTACGGGAGGAGCTAATAAACTTGCAGTAACTGGTGATATTTTCAATTAGGTTC 360
QY 382 ACTTTACAGGTACCGGTACTGTCAGCTACGGTGCAGTAACTGGTGATATTTTCAATTAGGTTC 441
D 361 ACTTTACAGGTACCGGTACTGTCAGCTACGGTGCAGTAACTGGTGATATTTTCAATTAGGTTC 420
QY 442 GGACAAGCTGCTGCTAATAATATACATATAGTCTTTTAGGTATATTAACCTAGTGGAGCG 501
D 421 GGACAAGCTGCTGCTAATAATATACATATAGTCTTTTAGGTATATTAACCTAGTGGAGCG 480
QY 502 AATGCCGGTTGACTATTGCTTCCAGATCCAGATGATTTAGGACCAATAACCTTGCAGGA 561
D 481 AATGCCGGTTGACTATTGCTTCCAGATCCAGATGATTTAGGACCAATAACCTTGCAGGA 540
QY 562 AATATAGTGGAGGAGGTATATAAATACCTGACAATAC 596
D 541 AATATAGTGGAGGAGGTATATAAATACCTGACAATAC 575

RESULT 7
RIOMPAB RIOMPAB 3087 bp DNA linear BCT 13-MAY-1996
LOCUS Rickettsia conorii outer membrane protein A (ompA) gene, partial
DEFINITION cds and repeat region.
ACCESSION L01462
VERSION L01462.1 GI:152485
KEYWORDS outer membrane protein A; repeat region.
SOURCE Rickettsia conorii (strain Kenya tick typhus) DNA.
ORGANISM Rickettsia conorii
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsia; Rickettsia; spotted fever group.
REFERENCE 1 (bases 1 to 3087)
AUTHORS Gilmore,R.D. Jr.
TITLE Comparison of the ompA gene repeat regions of Rickettsia reveals
species-specific arrangements of individual repeating units
JOURNAL Gene 125 (1), 97-102 (1993)
MEDLINE 93194085
PUBMED 7680636
FEATURES
source
1. .3087
/organism="Rickettsia conorii"
/strain="Kenya tick typhus"
/db_xref="taxon:781"
<1..>3087
/gene="ompA"
<1..>3087
/gene="ompA"
/codon_start=1
Location/Qualifiers
1. .3087
/organism="Rickettsia conorii"
/strain="Kenya tick typhus"
/db_xref="taxon:781"
<1..>3087
/gene="ompA"
<1..>3087
/gene="ompA"
/codon_start=1

```

Query Match	24.6%	Score 473.2	DB 1	Length 3087
Best Local Similarity	74.9%	Pred. No. 2.5e-77		
Matches 650	Conservative 0	Mismatches 203	Indels 15	Gaps 4
Qy 577	GGTATAA	TAAGTACATACAGATGCTGCCATTAACGGACAAATAGGTAATAC	TAAATCCG	636
Db 1912	GGTATAGTAACGGTTAACGGTGATAGTACAGTAAGTACGCAATATAGGTAATACAAATGCA	1971		
Qy 637	GCAGCTCAAAATTAAGCATTTGGAGCAAGTACGCTTCTTGAGGGGCGAGTTATTAAAGCC	696		
Db 1972	TTAGCAACAGTAATGTAGGACAGGTTTGCTACGAGTACAAGCGGAGTGGTAAATCG	2031		
Qy 697	ACTACGACTAAGTTTAAACAAATGCAGCGCGGTATTAAACACTTACAAATGCAAAATGCAGTA	756		
Db 2032	AATACAATAAACTTAACGGATAATGCGTCAAGGTGACATTTACGAATCC	2088		
Qy 757	TTACAGGTGCTGTTGATATAACACACAGCGCGGTGATGATGTAGGTGCTTTAAATTTAAAC	816		
Db 2089	GTGACCGGAGCGATAGATAAATACC	2145		
Qy 817	GGAGCGTTGACTCAAGTAACCTGGAATATAGTAAATACAAATTCATTAAGCGACAATAGT	876		
Db 2146	GG----	TAATAGTACAGTAACCTGGCAATATAGTAAATACAAATGCTATAGCAACAGTGAAT	2202	
Qy 877	GTAGGACGAGTACGGCTACGTTAGGGGGAGCGGTTATTAAAGCCACTACCACTAAAGTTA	936		
Db 2203	GTAGAGCAGGTATAGCCACCTTAGAGGGAGCGGTTATTAAAGCTACTACGACTAAATTA	2262		
Qy 937	ACGAATCGAGCGTCGGTATTAAACACTTAC	990		
Db 2263	ACGAATGCTGCGTATTAAACCTTACAAATGTAATGTCAGTATTAAACAGGTGCGATT	2322		
Qy 991	GATAACACACAGCGGTGATATGTAGTGTCGTCTAAATCTTAAGTGGAGCATTTGAGTCAA	1050		
Db 2323	GATAACACACAGCGGTTGATATGTAGTGTCGTCTAAATTTAAATTTAAATGGTGCATTAAGTCAA	2382		

```

Qy 1589 ACGGTAATAAGTCTACACTTAATATATATGCTGCTGCTACAGTGAATGCAAAATGATGTTA 1648
Db 121 ACGGTAATAAGTCTACACTTAATATATATGCTGCTGCTACAGTGAATGCAAAATGATGTTA 180
Qy 1649 GTATAGGCACAGTTCACAAATTAACATTCAAATTAATAAGATTTTGTATAAATGCTA 1708
Db 181 GTATAGGCACAGTTCACAAATTAACATTCAAATTAATAAGATTTTGTATAAATGCTA 240
Qy 1709 AGAAGCTGATGTTGATATATATAGACGCTCAGCGGATAGTTTAAAGGACGACTTCAC 1768
Db 241 AGAAGCTGATGTTGATATATATAGACGCTCAGCGGATAGTTTAAAGGACGACTTCAC 300
Qy 1769 GTCCTTTCTTACGTAACGTTAGTCTACAGATGATAGATTTATCACTCTTAAATAATCATTT 1828
Db 301 GTCCTTTCTTACGTAACGTTAGTCTACAGATGATAGATTTATCACTCTTAAATAATCATTT 360
Qy 1829 ACCCGGCTTCGTAACGCTGCTGCTGAGTAAATTCCTTTGAGTCCACAAAACCTATGAC 1888
Db 361 ACCCGGCTTCGTAACGCTGCTGCTGAGTAAATTCCTTTGAGTCCACAAAACCTATGAC 420
Qy 1889 CGTACAAGGTGACGTTGGAGCTCAAAAACATAG 1920
Db 421 CGTACAAGGTGACGTTGGAGCTCAAAAACATAG 452

RESULT 9
AF149108
LOCUS
DEFINITION
Rickettsia australis strain PHS outer membrane protein A (ompA)
gene, complete cds.
ACCESSION
AF149108
VERSION
AF149108.1 GI:5081770
KEYWORDS
Rickettsia australis.
ORGANISM
Rickettsia australis
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE
1 (bases 1 to 7067)
Stenos, J. and Walker, D. H.
The rickettsial outer-membrane protein A and B genes of Rickettsia
australis, the most divergent rickettsia of the spotted fever group
Int. J. Syst. Evol. Microbiol. 50 Pt 5, 1775-1779 (2000)
20487299
11034486
PUBMED
2 (bases 1 to 7067)
Stenos, J. and Walker, D.
Direct Submission
Submitted (10-MAY-1999), Australian Rickettsial Reference
Laboratory, Douglas Hocking Medical Institute, Geelong Hospital,
Ryrie St., Geelong, VIC 3220, Australia
FEATURES
source
1. 7067
/organism="Rickettsia australis"
/strain="PHS"
/db_xref="taxon:787"
354..6939
/gene="ompA"
354..359
/gene="ompA"
377..382
/gene="ompA"
409..413
/gene="ompA"
421..6741
/gene="ompA"
/codon_start=1
/transl_table=11
/product="outer membrane protein A"
/protein_id="AD39531.2"
/db_xref="GI:11641393"
/translation="MANISPKLFQKAIHKSALFTTSTAAMLSSSGAGWGAAGVI
SYNDAAFSNRAVANNWEITAGGAANGHADGPQDNEAFTYGGNHTITADEAGRIITA
INVAGTTPVALNSTQNTSVGSIVTGGNLLPVTIADGKSLITLGTAKVAADHGFDAAAD

```

repeat_region

/rpt_type=direct

/rpt_unit=1216..1470

6903..6939

/gene="ompA"

stem_loop

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 312;

Conservative

5; Mismatches

75; Indels

1; Gaps

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

1;

```
RESULT 10
RMU83439          3167 bp      DNA      linear      BCT 10-SEP-1998
LOCUS             Rickettsia mongolotimonae rOmpA (ompA) gene, partial cds.
DEFINITION        U83439
ACCESSION         U83439
VERSION           U83439.1  GI:1778884
KEYWORDS
SOURCE            Rickettsia mongolotimonae.
ORGANISM          Rickettsia mongolotimonae
                  Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
                  Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.
REFERENCE         1 (bases 1 to 3167)
AUTHORS           Fournier, P.E., Roux, V. and Raoult, D.
TITLE             Phylogenetic analysis of spotted fever group rickettsiae by study
                  of the outer surface protein rOmpA
JOURNAL           Int. J. Syst. Bacteriol. 48 Pt 3, 839-849 (1998)
MEDLINE           98404549
PUBMED            9734038
REFERENCE         2 (bases 1 to 3167)
AUTHORS           Raoult, D., Fournier, P.-E. and Roux, V.
TITLE             Direct Submission
JOURNAL           Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies,
                  CNRS BP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
FEATURES          Location/Qualifiers
                  1..3167
                     /organism="Rickettsia mongolotimonae"
                     /strain="HA-91"
                     /db_xref="taxon:45261"
                     1..3167
                        /gene="ompA"
                        /gene="ompA"
                        <1..>3167
                           /gene="ompA"
                           /codon_start=3
                           /transl_table=11
                           /product="rOmpA"
                           /protein_id="AAC35175.1"
                           /db_xref="GI:1778885"
                           /translation="ITLHAGSLDANNIDFGARSTLEFNGGGGIPYFKGAINGNNA
                  ILNVTLLTAYHITIGTVAEINAGNIEAIDASAGDVITLNAQDIFHRALDSTLL
                  SLNVTGVNIIAALDVLVAFGAEQGVVFNGLNIGSVNAGTARNIGDGNKFN
                  ALLIYNVATIDVNLQNLVNNADFTSFAFNAGAIQIONATYTIDANNGLN
                  IPAGNIOFAHADQILQNSGNDRTITLGANIDPDNDDEGIVILNSVTACKLITAG
                  KTFEGAKLQITVFKGDCSTAGTFTNTIVLITGQLELGAITANVLPKDAVQ
                  LTQNGTGGFLDFNAKGTIVNNVAVGTQVNTGNTGLIVLGASNLNRVNGIA
                  MLKVGAGNVTIAGKNVKGIEQGTNTLTPAHFNLGTSINKTGOALKLNFMSG
                  SVSGVGTAAANSVDGITTAGATSPASVNAKGTAKLGGTTSFAHTFTNTGAVTLAKGS
                  ITSFAKNVTATSFVANSATINFGSLAFNSITGSGTTLILGANQVYTGTSFTDL
                  TLNTTFDGAAGSGNLIKSGSTLDLSGVSTLALVVTATNFDNMNISPDKYTVISAE
                  TVGGLKPTPKENVKITLNDRFVDFTFDASTLPLEAKDIAADVIDKDFAPGGLANI
                  PNAANIKSLKLMEDANGSDARAFNNGLMTPLQEDATHTHLMQDVVRFSDITIAV
                  NQVVASINSITIALNARMKVQAGNKPVSSEDMDAKFGAWISPFVGNATQKMC
                  NSISGYSDTGTGTGTFDGFYSDDLVLGLAYTRADTDIKLNNKTGDKNKVESNIYSL
                  YGLXSVPEINLFEVIAIASYDNKIRSRVRIATLTETGVGTANGKYSKSYTGQLM
                  AGYTMMPENINLPLVCLRYSTIKDKGKETGYTQNLTVKGNYNTEFDGLLGAYS
                  SININVELTPELVAVDYAFKNKVSALDARLQGMTAPLPTNFKQSKTSFDFGVGV
                  TAKYMEYGINYDNTNGSKYFAQ"
BASE COUNT       1062 a 513 c 666 g 926 t
ORIGIN
Query Match      8.1%; Score 154.6; DB 1; Length 3167;
Best Local Similarity 69.8%; Pred. No. 9.4e-19;
Matches 224; Conservative 0; Mismatches 94; Indels 3; Gaps 1;
Qy 1469 GAATAACATTACGACCGGAGGACCTAGCTGCGAATAATATAGATTTTGGAGCTGCCA 1528
Db 1 GPATAACATTACGCTGGAGGAAGCTAGATGCGAATAATATAGATTTTGGAGCGCAGGA 60
Qy 1529 GTAATTTAGAGTTTAACTGCTCGCGGTGAGAAATTAACCTTAATTCGAACTATAGCAA 1588
Db 61 GTACTTTAGAGTTTAACTGCTCGCGGTGAGAAATTAACCTTAATTCGAACTATAGCAA 120
```

```
Qy 1589 ACGGTAATAATGCTACACTTAATTAATGCTGCTGTGCTACAGTTCGATTCGAAATGATGTA 1648
Db 121 ACGGCAATAATGCTATATTAATGTTAA--TACAAAGTCTACTACGSCATATCATTTAA 177
Qy 1649 GTATAGGCACAGTTCGCACAAATTAACATTCAAAATAATAAGATTTTCTTAATAATGCTA 1708
Db 178 CTATAGGAACAGTTCGCAGAAATCAATATTGGAGCGGTAAATATTTTTCGAATTGATGCAA 237
Qy 1709 AGAAGCTGATGTTCAATATATTAGACGCTCAGGCGATTTTAAAGAGCAGCGTCCAC 1768
Db 238 GTGCTGGTGATGTTACTATATTAATGCTCAAGATATTCAATTTAGAGCTCTAGATTCTA 297
Qy 1769 GTCTTTTCTTAGCTAACGTTA 1789
Db 298 CTTTATTACTTTCTAACTTAA 318
RESULT 11
AFI69629          629 bp      DNA      linear      BCT 19-JUL-2000
LOCUS             Rickettsia sp. FUJ98 outer membrane protein A gene, partial cds.
DEFINITION        AFI69629
ACCESSION         AFI69629
VERSION           AFI69629.1  GI:9280337
KEYWORDS
SOURCE            Rickettsia sp. FUJ98.
ORGANISM          Rickettsia sp. FUJ98
                  Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
                  Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.
REFERENCE         1 (bases 1 to 629)
AUTHORS           Zhang, P.H., Cao, W.C., Zhang, X.T., Xu, R.M., Dai, X.H. and Gao, D.Q.
TITLE             Identification of Spotted Fever Group Rickettsiae in Ticks from
                  Southern China
JOURNAL           Unpublished
REFERENCE         2 (bases 1 to 629)
AUTHORS           Zhang, P.H., Cao, W.C., Zhang, X.T., Xu, R.M., Dai, X.H. and Gao, D.Q.
TITLE             Direct Submission
JOURNAL           Submitted (16-JUL-1999) Epidemiology, Institute of Microbiology and
                  Epidemiology, 20 Dong-Da-Jie, Fengtai District, Beijing 100071,
                  P.R. China
FEATURES          Location/Qualifiers
                  1..629
                     /organism="Rickettsia sp. FUJ98"
                     /strain="FUJ98"
                     /db_xref="taxon:116094"
                     1..>629
                        /note="rOmpA"
                        /codon_start=1
                        /transl_table=11
                        /product="outer membrane protein A"
                        /protein_id="AAF86358.1"
                        /db_xref="GI:9280338"
                        /translation="MANISPKLFOKATQOGLKTAIFTTSTAAILMSSGVLGVAAGVI
                  AINDDAFSDNAENNMDEITAEVGANGIPAGGPONNNAFTSYADITITADVDRIT
                  ATNVDGTPVLGNIAQNTVGSITTRGNLPTVITAGKSLTNGLNNAVNAAGFDAPAD
                  NYTGIGNITLGGANAELIQSATPAKTIAGNIDGGIITVKTDAAINGT"
BASE COUNT       190 a 113 c 143 g 183 t
ORIGIN
Query Match      7.9%; Score 151.2; DB 1; Length 629;
Best Local Similarity 56.4%; Pred. No. 5.9e-18;
Matches 360; Conservative 0; Mismatches 248; Indels 30; Gaps 3;
Qy 1 ATGGCGAATATTTCTTAAATTTTCAAAAAGCAATTTCAAAAAGGCTCTTAAACTGCT 60
Db 1 ATGGCGAATATTTCTTAAATTTTCAAAAAGCAATTTTCAAAAAGGCTCTTAAACTGCT 60
Qy 61 TTATTCCACCCTCAACCGCAGCGATATGCTTAACCGGTAGTGAGTCTCTTGGTGTGCA 120
Db 61 TTATTCCACCCTCAACCGCAGCGATATGCTAGTAGTAGTGGGTACTCGGTGTGCT 120
Qy 121 AGAACCGTAACGTGATGGTGCAGAGCTTGCAGCCGGAACAAATATAGTCTCTGAGGCC 180
Db 121 GCAGGTGTTATGCTTACTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
```


Qy	181	GGTGTCTTTTGTAGCGGGTCTACTTTTACAATATACCGGTGCTTTTACGGTTACTGATGCT	240
Db	181	TGGGATGAGATAACGGGTGAAGGGTAGCTAATGGTATTCTCTGGCGGTCTCAAAAC	240
Qy	241	GACCTAAGTCTTCGTGCATAGATTAGATTAAATAATTTTGCACAGGCTCT	287
Db	241	AATTGGGCATTTTACTTTACAGTGTCTGATTATACATATCACTGCAGATGTAGTCGATCGTATT	300
Qy	288	-----TTTTTCAGTAACTGCTGTATATTTCAATAGTTTCAGTGGTAGATACGGGAGGA	339
Db	301	ATTAGGCGCTAAATGTGCGGCTACTACTCCGTAGGTCTAAATATTCTCANAATACC	360
Qy	340	GCTAATAAATCTTCGAGTTAATATTGATGATGGTTTAACTTTACAGGTACCGGT	399
Db	361	GTGCTTGGTTTCGATTATACTAGAGTAACTTTGTCCCTGTTACTATTCTCGCAAAAGC	420
Qy	400	ACTGCAAGCCTACGGGTCAAAATCTCGTGTATTTCCAAGGTGGACAAGCTGCTGCTAAT	459
Db	421	TTAACTTTAAACGGTACTAATGCTGTGCTGCNAATCATGGTTTTCATGCTCCGCCGAT	480
Qy	460	AATACATATACTGCTTTAGGTAAATATACTCTAGGTGGAGCGAATGCCGGTTTGACTATT	519
Db	481	AAT---TATACAGGTTTAGGAAATAAATCTTTAGGGGGAGCGAATGCTGAATTAATATA	537
Qy	520	GCATTACAGATCCAGATGTATTTAGSACCAATAAACGCTTTGCAGGAAATATAGATCGAGGAGT	579
Db	538	CAATCTGCACCCCGCGAAG-----ATAACACTTGCAGGAATATAGATGGAGAGGT	591
Qy	580	ATAATACTGACAATACAGATGCTGCCATTAACGGAAAC	617
Db	592	ATAATACTGCTCAAGACAGATGCTGCCATTAACGGAAAC	629

RESULT	12
AF179366	
LOCUS	3168 bp DNA linear BCT 12-SEP-2000
DEFINITION	Rickettsia sp. HL-93 strain HL-93 ompA (ompA) gene, partial cds.
ACCESSION	AF179366
VERSION	AF179366.1 GI:9789177
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	Zhang,J.Z., Fan,M.Y., Wu,Y.M., Pournier,P.E., Roux,V. and Raoult,D.
TITLE	Genetic classification of 'Rickettsia heilongjiangii' and 'Rickettsia hulunui', two Chinese spotted fever group rickettsiae J. Clin. Microbiol. 38 (9), 3498-3501 (2000)
JOURNAL	
MEDLINE	20440687
PUBMED	10970415
REFERENCE	2 (bases 1 to 3168)
AUTHORS	Zhang,J.Z. and Raoult,D.
TITLE	Direct Submission
JOURNAL	Submitted (18-AUG-1999) Unite des Rickettsies, Faculte de Medecine,
FEATURES	27 Bd. Jean Moulin, Marseille 13385, France
Source	Location/Qualifiers 1..3168 /organism="Rickettsia hulunui" /strain="HL-93" /db_xref="taxon:121193" <1..>3168 /gene="ompA" <1..>3168 /genes="ompA" /note="rOmpA" /codon_start=3 /transl_table=11 /product="OmpA" /protein_id="AAF98268.1" /db_xref="GI:9789178" /translation="LHVTGGSLDANNIDFGARSTLFNCGGCIYYFKGAINGNNNA"
gene	
CDS	

RESULT 14
RSU83455

QY 1640 ATGATGTTAGTATAGGCACAGTTCACAAATTAACATTAACAAATAAAGATTTTGTAA 1699
 Db 178 ATCAATTAACATAGACAGTTCAGAAATCAATATTGGAGCGGTAATATTTTGCAG 237
 QY 1700 TAAATGCTAAGACGCTGATGTTGATATATTAGACGCTCAGCGATAGTTTAAAGAG 1759
 Db 238 TTGATGCAAGTGCTGGTGTACTATATTAAATGCTCAAGATATTATTTTAGAGCTC 297
 QY 1760 CAGCTTACAGCTTTTCTTAGCTAACGTTAGTCTACAGATGATAGAGTTATCACTCTTA- 1818
 Db 298 TAGATTCTACTTACTTCTTAACCTTAACCGGAGTCGGAGTAAATATATATTATAG 357
 QY 1819 AAATCATTTACCGGCTTCTGCTAACGGTGGTGGTGAATTTTGTAGTCCACAA 1878
 Db 358 CAGCTGATTAGTACTCCCGTGTGCTGAAGGTACGCTTAATCTTTAATGGTGGGTTA 417
 QY 1879 AACTTATCAGCTCAAGCTGACGTTGAGCTA 1911
 Db 418 ATGGTCTGCATATTGGGAGTAATGTAGCAGGTA 450

RESULT 15
 RSU83438
 LOCUS RSU83438 3188 bp DNA linear BCT 10-SEP-1998
 DEFINITION Rickettsia BAR-29 rOmpA (ompA) gene, partial cds.
 ACCESSION U83438
 VERSION U83438.1 GI:1778898
 KEYWORDS
 SOURCE Rickettsia sp. Bar29.
 ORGANISM Rickettsia sp. Bar29.
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
 Fournier, P.E., Roux, V. and Raoult, D.
 Phylogenetic analysis of spotted fever group rickettsiae by study
 of the outer surface protein rOmpA
 Int. J. Syst. Bacteriol. 48 Pt 3, 839-849 (1998)
 98404549
 9734038
 2 (bases 1 to 3188)
 Raoult, D., Fournier, P.-E. and Roux, V.
 Direct Submission
 Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies,
 CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
 Location/Qualifiers
 1..3188
 /organism="Rickettsia sp. Bar29"
 /strain="Bar 29"
 /db_xref="taxon:45260"
 1..3188
 /gene="ompA"
 <1..>3188
 /gene="ompA"
 /codon_start=3
 /transl_table=11
 /product="rOmpA"
 /protein_id="AAC35174.1"
 /db_xref="GI:1778898"
 /translation="ITIQAGGSLLAANNIDFGVSTLEFNGPLDGGNTIPYFKGAIA
 NGNAILNVNKSITAYHIGTVAEINIGAGNFAIDASADDVITLNAQAINFGVDP
 SALVSLNLTVEVKILLAADLVAPGANGYVYVNGVNGVNGVNGVNGVNGVNGVNGV
 GDFNTLLLYNAVITIDDDVNELEGIONHINNAAFTSSAFNAGAIQINDATYIIDAN
 NGNLNPAGNIQFAHNAQLIQNTSGNDRTITLGANIDPDNDDEGIVTLNSVTAGK
 LDTAGKMLGGAAHKLHITVFKAGDGAAGTFTNTNIVLDTTGLRELGATTANVLL
 NDAIQDTGGTNGTDFDNKNGTTLNNVNVAGVQNTGGTNGTNGTILVGSNLN
 SVNGIAMLKVGAGNVTKIGKIGIOGTGNTLTLPANFLTGSINKTGGQALKL
 NPTNGSVGVCTAANLVGDITTAGATSEVSSVNAKGTATLGTTSEADTFPTNGAV
 TLAKGSITNFAKNVATSFVANSATINFGNSLAFNSNITGSGTTLILGANQVYTAGT
 GSTFDLTNTFDGAKSGGNILIKSGTLDLSGVSTLALVVVATNFNDINNISPDK
 YTVISATPGGLPTPEENVKTIINNDNHVDFPTFDSATLTFADIAADVIDEAFEP
 GGPLANIPNAIKKLELMENAPNGSDARFNFNGLMPTLQEADATTHLMQDVVKP
 SDTIAAVNQVNASNITSSITARNMRKVQAVNKGPSVSGDEMDAKFGAWISPFVG
 NATQKMCNNISGVKSDTTGTGTGFDGLVSDLLALGLAYTRADTDIKLKNKTKDKNKY

ESNTSYLYGLYNVPEYENLFVEAIASYSDNKISSKSRVIAETALETGYQIANGKYKSE
 SCTGQLMVGTYTMMPEINILTPLAGLRYSTIKDKYKGTCTTQNLTVKGNKNTEDG
 LLGAKVSSNNVNEILVLTPELYAMVDYAFKNKVPADARLQGHFAPLPTNSFQSKSTS
 FDVGVGVTAKHKMWEYGINIDTINIGSKYFAQQ

BASE COUNT 1061 a 539 c 670 g 918 t
 ORIGIN

Query Match 7.18; Score 136; DB 1; Length 3188;
 Best Local Similarity 67.98; Pred. No. 2.5e-15;
 Matches 228; Conservative 0; Mismatches 90; Indels 18; Gaps 2;

QY 1469 GAATAACATTACGACCGGAGGAGCCTAGCTGCGAATAATATAGATTTTGGAGCTGCGA 1528
 Db 1 GAATAACATTACAAAGCTGGAGGAGCCTAGCTGCGAATAATATAGATTTTGGAGCTGCGA 60
 QY 1529 GTAATTAGAGTTTAAACGGTCTGCCGGT-----AAGAAATTATAACTTAA 1573
 Db 61 GTACTTTAGAGTTTAAACGGACCTCTTGATGGTGGTAAACACAATTCCTTTATTATTATA 120
 QY 1574 TCGGAACCTATAGCAACGGGTAATAATGCTACACTTAATATTAATGCTGCTACAGTGA 1633
 Db 121 AAGGAGCTATAGCAACGGCAATAACGCTATATTAATGTTAA---TACAAAGTCACTTA 177
 QY 1634 TTGCAAAATGATGTTAGTATAGGCACAGTTGCACAAATTAACATTCACAAATAAAGATTT 1693
 Db 178 CGGCATATCATTCACACTATAGGAACAGTTGCAGAAATTAATATTGGAGCCGGTAATTTT 237
 QY 1694 TTGTAATAAATGCTAAGAACGCTGATGTTGATATATTAGACGCTCAGCGGATTAGTTTAA 1753
 Db 238 TTGCAATTGATGCAAGTCTGATGATGTTACTATATTAAATGCTCAAGCTATTAAATTTG 297
 QY 1754 AAGGAGCAGCTTACGCTCTTTCTTAGCTTAACAGTTA 1789
 Db 298 GAGTTCAGATTCTGCTTTAGTACTTTTCTTAACTTAA 333

Search completed: April 17, 2003, 12:04:11
 Job time : 9824 secs

XX WPI; 2001-596829/67.
 DR P-PSDB; AAE10890.
 XX
 XX Novel isolated Rickettsia felis outer membrane protein and
 PT polynucleotide for modulating expression of the protein in a host cell
 PT
 XX
 PS Example 1; Page 64-68; 73pp; English.
 XX
 CC The present invention relates to an isolated Rickettsia felis outer
 CC membrane protein (rompA) and its polynucleotide. The rompA gene is
 CC useful for modulating expression of the protein in a host cell. The
 CC rompA gene is useful as probe or for the design of primers to obtain
 CC DNA encoding the protein by either cloning and colony/plaque
 CC hybridisation or amplification using PCR. The present sequence is
 CC Rickettsia felis outer membrane protein genomic DNA (rompA).
 XX
 SQ Sequence 5513 BP; 1814 A; 961 C; 1181 G; 1557 T; 0 other;
 Query Match 100.0%; Score 1920; DB 22; Length 5513;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGGAATATTTCTTAAATTTTCAAAAAGCAATTCAAAAGGCTTTAAACTGCT 60
 DB 478 ATGCGGAATATTTCTTAAATTTTCAAAAAGCAATTCAAAAGGCTTTAAACTGCT 537
 QY 61 TTATTACCACCTCAACCGCAGCGATATGCTAACCGGTAGTGAGTCCCTGGTGCA 120
 DB 538 TTATTACCACCTCAACCGCAGCGATATGCTAACCGGTAGTGAGTCCCTGGTGCA 597
 QY 121 AGAACCGTAAGTGTGATGTGTCAGAGCTTGCAGCCGGAACAATATAGGTCCTGGAGCC 180
 DB 598 AGAACCGTAAGTGTGATGTGTCAGAGCTTGCAGCCGGAACAATATAGGTCCTGGAGCC 657
 QY 181 GGTCTTTTGTAGCGGCTTCTACTTTTCAATATATACCGGTGCTTTACGGTTACTGATGCT 240
 DB 658 GGTCTTTTGTAGCGGCTTCTACTTTTCAATATATACCGGTGCTTTACGGTTACTGATGCT 717
 QY 241 GACGTAAGTGTGCTGATTTAGATTTTAAATTTTTCAGCAGGCTCTTTTTCAGTAAC 300
 DB 718 GACGTAAGTGTGCTGATTTAGATTTTAAATTTTTCAGCAGGCTCTTTTTCAGTAAC 777
 QY 301 GGTGATATTTCAATAGTGTGATGTCAGGAGAGCTAATAAAGTTCAGTTAAT 360
 DB 778 GGTGATATTTCAATAGTGTGATGTCAGGAGAGCTAATAAAGTTCAGTTAAT 837
 QY 361 ATTGATGATGTTTAACTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACT 420
 DB 838 ATTGATGATGTTTAACTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACT 897
 QY 421 CCTGCGTTGTTATCCAAAGTGGCAAGCTGCTGCTAATAATACATATACATGCTTTAGGT 480
 DB 898 CCTGCGTTGTTATCCAAAGTGGCAAGCTGCTGCTAATAATACATATACATGCTTTAGGT 957
 QY 481 AATATACTCTAGGTGAGCGAATGCCGGTTTGACTATTGCTTCAGATCAGATGATTA 540
 DB 958 AATATACTCTAGGTGAGCGAATGCCGGTTTGACTATTGCTTCAGATCAGATGATTA 1017
 QY 541 GGACCAATAACGCTTGAGGAAATATAGATGGAGGAGGTATATACTGACATACAGAT 600
 DB 1018 GGACCAATAACGCTTGAGGAAATATAGATGGAGGAGGTATATACTGACATACAGAT 1077
 QY 601 GCTGCCAATTAACGGAACAATAGGTATATCTAATCCCGCAGCTCAATTAAGCATTTGGAGCA 660
 DB 1078 GCTGCCAATTAACGGAACAATAGGTATATCTAATCCCGCAGCTCAATTAAGCATTTGGAGCA 1137
 QY 661 AGTACGCTTTCTCTTGAGGGGCGAGTTATTAAAGCCACTACGACTAGTTTAAACAATGCA 720
 DB 1138 AGTACGCTTTCTCTTGAGGGGCGAGTTATTAAAGCCACTACGACTAGTTTAAACAATGCA 1197
 QY 721 GCGCCGGTATTAACACTTTACAAATGCAATGCAGTATTAAACAGGTGCTGTTGATAACACC 780

DB 1198 GCGCGGTATTACACATTACAAATGCAATGCACTATTAAACAGTCTGTTGATAACACC 1257
 QY 781 ACAGCGCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 DB 1258 ACAGCGCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1317
 QY 841 AATATAGTATACAAATTCATTAGCGACATAAGTGTAGGAGCAGGTACGGCTACGTTA 900
 DB 1318 AATATAGTATACAAATTCATTAGCGACATAAGTGTAGGAGCAGGTACGGCTACGTTA 1377
 QY 901 GGGGAGCGGTATTAAAGCCACTACGACTAAAGTAACTAACTAACTAACTAACTAACTAACTAA 960
 DB 1378 GGGGAGCGGTATTAAAGCCACTACGACTAAAGTAACTAACTAACTAACTAACTAACTAACTAA 1437
 QY 961 CTTACAATGCGAGTATTAAACAGGTGCTGTGATTAACACACAGCGCGGTGATATGATAGT 1020
 DB 1438 CTTACAATGCGAGTATTAAACAGGTGCTGTGATTAACACACAGCGCGGTGATATGATAGT 1497
 QY 1021 GTCGTAATCTAAGTGGAGCATTGAGTCAAGTAACTAACTAACTAACTAACTAACTAACTAACTAA 1080
 DB 1498 GTCGTAATCTAAGTGGAGCATTGAGTCAAGTAACTAACTAACTAACTAACTAACTAACTAACTAA 1557
 QY 1081 TTAGCCACGATAATATAGGAGCAGGTGTGCTACCTTGGATGGAGCGGTATTATAAAGCT 1140
 DB 1558 TTAGCCACGATAATATAGGAGCAGGTGTGCTACCTTGGATGGAGCGGTATTATAAAGCT 1617
 QY 1141 ACTACGACTAAGTTAACAGATGATGCTGCTAGTATTATATTTACGATCTGTAGTAGTA 1200
 DB 1618 ACTACGACTAAGTTAACAGATGATGCTGCTAGTATTATATTTACGATCTGTAGTAGTA 1677
 QY 1201 ACCGGACAATAGATAATACCGGTAACTGCAATAAAGGTGTGTTAACTTTACCGGAGCA 1260
 DB 1678 ACCGGACAATAGATAATACCGGTAACTGCAATAAAGGTGTGTTAACTTTACCGGAGCA 1737
 QY 1261 AGTACGTAACCGATAATATAGGTAACTGCAATAAAGGTGTGTTAACTTTACCGGAGCA 1320
 DB 1738 AGTACGTAACCGATAATATAGGTAACTGCAATAAAGGTGTGTTAACTTTACCGGAGCA 1797
 QY 1321 GGTGCTGCTGCAATACAAAGCGGAGTAGTAAAGCGAATGCAATAACTTTAACTTTAACTTTAACT 1380
 DB 1798 GGTGCTGCTGCAATACAAAGCGGAGTAGTAAAGCGAATGCAATAACTTTAACTTTAACTTTAACT 1857
 QY 1381 GCGTACGTAGTACATTTACCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 DB 1858 GCGTACGTAGTACATTTACCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1917
 QY 1441 CTCTTCGCAACAGTGAATATAGGAGCAGGAATACATTTACGAGCGGAGGAGCGCTAGCT 1500
 DB 1918 CTCTTCGCAACAGTGAATATAGGAGCAGGAATACATTTACGAGCGGAGGAGCGCTAGCT 1977
 QY 1501 GCGAATAATATAGATTTTGGAGCTGGAGTAACTTTAGAGTTTAACTGCTCTGCGGTAAG 1560
 DB 1978 GCGAATAATATAGATTTTGGAGCTGGAGTAACTTTAGAGTTTAACTGCTCTGCGGTAAG 2037
 QY 1561 AATTAACCTTAATCGAATATAGCAACCGGTAACTGCTACACTTAATATTAATGCT 1620
 DB 2038 AATTAACCTTAATCGAATATAGCAACCGGTAACTGCTACACTTAATATTAATGCT 2097
 QY 1621 GCTGGTACAGTGTGCAATGATGTTAGTATAGGACAGCTTGCACAAATTAACATTCAA 1680
 DB 2098 GCTGGTACAGTGTGCAATGATGTTAGTATAGGACAGCTTGCACAAATTAACATTCAA 2157
 QY 1681 AATAAAGATTTTGTATTAATGCTTAAGACGCTGATGTTGATATATTAGACGCTCAG 1740
 DB 2158 AATAAAGATTTTGTATTAATGCTTAAGACGCTGATGTTGATATATTAGACGCTCAG 2217
 QY 1741 GCGATAGTTTAAAGGAGCAGCTTCACTGCTTTTCTTAGCTAACGTTAGCTACAGATG 1800
 DB 2218 GCGATAGTTTAAAGGAGCAGCTTCACTGCTTTTCTTAGCTAACGTTAGCTACAGATG 2277
 QY 1801 ATAGAGTTATCACTCTTAAATCAATTTACCGGCTTGTGTAACGGTGGTGGTGAATTA 1860

Db 2278 ATAGATTATCACTCTTAAATAATCATTTACCGGCTTGTCTAACGGTGGTGGAGTTAA 2337
 QY 1861 TTCTTTTGGTCCAAACAACTATTGACCGGTACAGGTGACGTGAGCTAAACAATAG 1920
 Db 2338 TTCTTTTGGTCCAAACAACTATTGACCGGTACAGGTGACGTGAGCTAAACAATAG 2397

RESULT 2
 AAF75507/c
 ID AAF75507 standard; DNA; 486 BP.
 AC AAF75507;
 XX
 XX
 DT 09-MAY-2001 (first entry)
 XX
 XX Polyglutamine tract coding sequence #2.
 XX
 KW Animal model; polyglutamine tract; neurodegenerative disorder; HDJ1;
 KW heat shock protein 40; tetrairicopeptide repeat protein 2; tpr2; stroke;
 KW myeloid leukaemia factor 1; MLF; human; fruit fly; Alzheimer's disease;
 KW Parkinson's disease; CJD; BSE; Huntington's disease; head trauma;
 KW cancer; ds.
 XX
 OS Drosophila sp.
 XX
 XX WO200112238-A1.
 PN
 XX
 PD 22-FEB-2001.
 XX
 PF 14-AUG-2000; 2000WO-US22496.
 XX
 PR 12-AUG-1999; 99US-0148933.
 PR 12-AUG-1999; 99US-0148934.
 PR 18-JAN-2000; 2000US-0177047.
 PR 19-MAY-2000; 2000US-0205720.
 XX
 XX (CALY) CALIFORNIA INST OF TECHNOLOGY.
 PA
 XX Benzer S, Kazemi-Esfarjani P;
 PI
 XX WPI; 2001-147537/15.
 DR
 XX
 XX Identifying genes or other compounds that modulate polyglutamine
 PT toxicity, useful for treating Alzheimer's disease, Parkinson's disease
 PT and Creutzfeldt-Jakob disease
 XX
 PS Disclosure; Fig 1A; 275pp; English.
 XX
 CC The present invention describes a method of screening for genes which
 CC modulate polyglutamine toxicity using animal models with polyglutamine
 CC sequences that cause toxicity in the animal. The model is preferably
 CC Drosophila, and toxic polyglutamine sequences include the human and
 CC Drosophila heat shock protein 40/HDJ1, tetrairicopeptide repeat protein 2
 CC (TPR2) and myeloid leukaemia factor 1 (MLF) genes. The model is useful
 CC for identifying treatments for neurodegenerative and proliferative
 CC disorders, including Alzheimer's disease, Parkinson's disease,
 CC Creutzfeldt-Jacob's disease (CJD), bovine spongiform encephalopathy
 CC (BSE), Huntington's disease, Machado-Joseph disease, spinocerebellar
 CC ataxias, dentatorubropallidolysian atrophy, Kennedy's disease, stroke,
 CC head trauma and cancer.
 XX
 SQ Sequence 486 BP; 193 A; 164 C; 108 G; 21 T; 0 other;

Query Match 3.3%; Score 63.8; DB 22; Length 486;
 Best Local Similarity 48.2%; Pred. No. 4e-06;
 Matches 179; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 110 TTGGTGCTGCAAGAACCGTAACTGATGTCGACAGCTTGCAGCCGCAACAATATAG 169
 Db 441 TTGCTGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 382
 QY 170 GTCCGTGAGCCGGTCTTTTGTAGCGGGTTCACATTTACAATATACCGGTGTTTACGG 229

Db 381 TTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 332
 QY 230 TTACTGATGCTGACGCTAGCTGTTGCTGCAATAGATTTAATAATTTTGACAGAGCTCTT 289
 Db 321 CTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 262
 QY 290 TTTCAGTAACTGCTGATATTTTCATTAGGTTTCAGTGTAGATACGGGAGGAGCTAATAAAC 349
 Db 261 CTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 202
 QY 350 TTGCACTTAATATGATGATGTTTAACTTAACCTTAACTTAACTGACGCTACGACGCT 409
 Db 201 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 142
 QY 410 ACGGTGCAATCCTGCTGTTGTTATTCCAAAGTGGACAAGCTGCTGCTAATAATACATA 469
 Db 141 TTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 82
 QY 470 CTGCTTTAGGT 480
 Db 81 CTGCTGTTGCT 71

RESULT 3
 ABA48451/c
 ID ABA48451 standard; DNA; 510 BP.
 XX
 AC ABA48451;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human breast cell single exon nucleic acid probe #7146.
 KW Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-496933/54.
 XX
 PT New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes
 XX
 PS Claim 4; SEQ ID NO 7146; 327pp + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene

XX Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 OS Homo sapiens.
 XX WO200157274-A2.
 XX 09-AUG-2001.
 PF 30-JAN-2001; 2001WO-US00666.
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488899/53.
 DR Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 XX Claim 4; SEQ ID NO 11883; 530pp; English.
 XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 510 BP; 149 A; 184 C; 68 G; 109 T; 0 other;
 SQ

Query Match 3.1%; Score 59.2; DB 22; Length 510;
 Best Local Similarity 46.1%; Pred. No. 5.8e-05;
 Matches 199; Conservative 0; Mismatches 233; Indels 0; Gaps 0;
 QY 104 GAGTCCTTGGTGTGCAAGAACCGTAACCTGCTGATGCTGAGCTGCGAGCGGAAACAA 163
 DB 510 GTGTTGATGCTGATGAGTACCGATGCTGGAAGCAGGCTGTGATGAGCTGGAAGAAAG 451
 QY 164 ATATAGTCTCTGAGCGGCTGCTTTTGTAGCGGGTCTACTTTACAAATATACCGGTCCT 223
 DB 450 GTGTTGTCAGGTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 391
 QY 224 TTACGGTTTACTGATGCTGACGTAAGTGTTCGTGCAATAGATTAAATATTTTCAGCAG 283
 DB 390 ATGACAGTATTGCTGTTGTCAGATGTTGATGTCAGATGTTGATGTCAGATGTTGATGTC 331
 QY 284 GTCCTTTTTCAGTAACCTGATATTTTATAGTTCAGTGTGATAGTACGGAGGAGCTA 343
 DB 330 GTGTTGATGAGTGCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 271
 QY 344 ATAACTTTCAGTAAATATTGATGATGTTTAACTTTAACTTTAACTTTAACTTTAACTTTAA 403
 DB 270 CAGATGTTGATGATGATGTTGAGCAGGTTGATGATGATGATGATGATGATGATGATGATG 211
 QY 404 CAGCCTACCGTGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 463
 DB 111111111111111111111111111111111111111111111111111111111111

Db 210 CAAGTGCAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 151
 QY 464 CATATACTCTTTAGTAAATATAAATCTAGGTGGACGGAATGCCGCTTTGACTATTGCTT 523
 Db 150 TTGATGTCAGATGTTGAAGCGGGTGTGATGCTGATGCTGATGCTGCTGCTGCTGCTGCTG 91
 QY 524 CAGATCCAGATG 535
 Db 90 CTGATACAGATG 79
 RESULT 6
 AAK14772/c
 ID AAK14772 standard; DNA; 510 BP.
 XX AAK14772;
 XX 05-NOV-2001 (first entry)
 DT Human brain expressed single exon probe SEQ ID NO: 14763.
 DE Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer; ss.
 XX Homo sapiens.
 OS WO200157275-A2.
 PN 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00667.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-483446/52.
 DR Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX Example 4; SEQ ID NO: 14763; 650pp + Sequence Listing; English.
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.
 XX Sequence 510 BP; 149 A; 184 C; 68 G; 109 T; 0 other;
 SQ

Query Match 3.1%; Score 59.2; DB 22; Length 510;
 Best Local Similarity 46.1%; Pred. No. 5.8e-05;
 Matches 199; Conservative 0; Mismatches 233; Indels 0; Gaps 0;
 QY 104 GAGTCCTTGGTGTGCAAGAACCGTAACCTGCTGATGCTGAGCTGCGAGCGGAAACAA 163
 DB 510 GTGTTGATGCTGATGAGTACCGATGCTGGAAGCAGGCTGTGATGAGCTGGAAGAAAG 451
 QY 164 ATATAGTCTCTGAGCGGCTGCTTTTGTAGCGGGTCTACTTTACAAATATACCGGTCCT 223
 DB 450 GTGTTGTCAGGTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 391

Qy	224	TTACGGTTACTGATGCTGACGTAAAGTTCGTGTCATAGATTAAATAATTTTGACGACG	283
Db	390	ATGCAGGTATTGGTGTGGTGACAGATTGTGATSCAGATGTTGAAGCAGGTGCTGGAANA	331
Qy	284	GTCTTTTTTCAGTAACCTGGTGATATTTCAATTAGTTTCACTGGTGAGATACGGAGAGACTA	343
Db	330	GTGTGTGATGCAGGTGCAGATGCTGATGCTGCTGGTAGTGATGCTGATGGTAGTGCTTTGG	271
Qy	344	ATAAACCCTTGCAGTAAATATTGATGATGGTTTAACTTAACTTAAACAGTACCGGTACTG	403
Db	270	CAGATGTTGTATGCAGATGTTTGAACGACAGGTGTGATCAGGTGCTGAAAAATGTTTGATG	211
Qy	404	CAGCCTACGGTGAATCCCTGCGTGTGTTATCCAAAGGTGCACAAAGCTGCTGCTAATAATA	463
Db	210	CAAGTCAGATGCTGATGCTGGTAGTGATGCTGATGCTGGTAGTGATGCTGGTGCGAGATG	151
Qy	464	CATATACTGCTTTAGGTAAATATAACTCTAGGTGGACGAATCCGGTTTGAATTTGCTT	523
Db	150	TTGATGCAGATGTTGAAGCGGGTGTGATCCTGATCCAGGTGCTGAACGAGGTATTGGTG	91
Qy	524	CAGATCCAGATG	535
Db	90	CTGATACAGATG	79

RESULT 7	
AAK40512/c	
ID	AAK40512 standard; DNA; 510 BP.
XX	
XX	
AC	AAK40512;
XX	
XX	
DT	06-NOV-2001 (first entry)
XX	
XX	
DE	Human bone marrow expressed single exon probe SEQ ID NO: 15069.
XX	
XX	Human; bone marrow expressed exon; gene expression analysis; probe;
KW	microarray; cancer; leukaemia; lymphoma; myeloma; ss.

AA	Homo sapiens.
OS	
XX	
XX	WQ200157276-A2.
PN	
XX	
XX	
PD	09-AUG-2001.
XX	
XX	
PF	30-JAN-2001; 2001WO-US00668.
XX	
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	03-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
XX	(MOLE-) MOLECULAR DYNAMICS INC.
PA	
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
XX	WPI; 2001-488900/53.
XX	
PT	Human genome-derived single exon nucleic acid probes useful for
PT	analyzing gene expression in human bone marrow -
XX	
XX	
PS	Example 4; SEQ ID NO: 15069; 658pp + Sequence Listing; English.
XX	
XX	The present invention provides a number of single exon nucleic acid
CC	probes which are derived from genomic sequences expressed in the human
CC	bone marrow. They can be used to measure gene expression in bone marrow
CC	samples, which may enable the improved diagnosis and treatment of cancers
CC	such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC	the probes of the invention.
XX	
XX	

Sequence	510 BP; 149 A; 184 C; 68 G; 109 T; 0 other;
Query Match	3.1%; Score 59.2; DB 22; Length 510;
Best Local Similarity	46.1%; Pred. No. 5.8e-05;
Matches	199; Conservative 0; Mismatches 233; Indels 0; Gaps
Qy	104 GAGTCCCTGGTCGCAAGAACCGTTAACTCCTCATGTGTCGACAGCTTTGCAGCGGGAACA 163
Db	510 GTGTTGATGCTGATGCAAGTGACCGATGCTGAAGCAGGTGTTGATGCAGGTGCTGAAAAAG 451
Qy	164 ATATAGTCTCGGAGCCGGTCTTTTTGTACGCCGGTTCTACTTTTACAAATATACCGGTGCTT 223
Db	450 GTGTTGTCGAGGTGCAGATGCTGATGCTGGTAGTAATGCTGATGCAGCGCTGATGCTG 391
Qy	224 TTACGGTTACTGATGCTGCAGCTAAGTGTTCGTGCATTAGATTTAAATAATTTTGCACGAC 283
Db	390 ATGCAGCTATTGGTGTGGTGCAGATGCTTGATGCAGATGTTGAAGCAGGTGCTGAAAAAG 331
Qy	284 GTCITTTTTTCAGTAACHTGGTGATATTTCAATTAGGTTTCAGTGGTAGATACGGGAGGAGCTA 343
Db	330 GTGTTGATGCAGGTGCAGATGCTGATGCTGTTAGTATGCTGATGCTGATGCTGATGCTGTTG 271
Qy	344 ATAACCTTGCAGTTTAATTATGATGGTTTAACTTTAACTTTAAACAGGTACCGGTACTG 403
Db	270 CAGATGTTGATGCAGATGTTGAAGCAGGTGTTGATGCAGGTGCTGAAAATGTGTTGATG 211
Qy	404 CAGCCTACGGTGCNAAPFCCTCGCTGTTATTCACAGGTGGACAAAGTCTGCTTAATAATA 463
Db	210 CAAGTGCAGATGCTGATGCTGCTGGTAGTAGTCTCATGCTGCTAGTATGCTGGTGCAGATG 151
Qy	464 CATATACTGCTTTAGGTAAATAACATCTAGGTGGAGCGAATGCCGGTTTGCATTAATTCGCTT 523
Db	150 TTGATGCAGATGTTGAAGCCGGTGTGTTGATGCTCATGAGGTGCTGAAGCAGGTATTGGTG 91
Qy	524 CAGATCCAGATG 535
Db	90 CTGATACAGATG 79
RESULT 8	
AAI46546/c	
ID	AAI46546 standard; DNA; 510 BP.
XX	
AC	AAI46546;
XX	
DT	17-OCT-2001 (first entry)

RESULT 8	
AAI46546/c	
ID	AAI46546 standard; DNA; 510 BP.
XX	
XX	AAI46546;
AC	
XX	
DT	17-OCT-2001 (first entry)
XX	
DE	Probe #15232 used to measure gene expression in human placenta sample
XX	
KW	Probe; microarray; human; placenta; antenatal diagnosis;
KW	genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200157272-A2.
XX	
PD	09-AUG-2001.
PF	
XX	30-JAN-2001; 2001WO-US00663.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WPI; 2001-488897/53.

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 12:04:21 ; Search time 193 Seconds
(without alignments)
3050.878 Million cell updates/sec

Title: US-09-800-065-1

Perfect score: 1920

Sequence: 1 atggcggaattttctctaaa.....cgttgagctaaacaatag 1920

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	60.4	3.1	7218	1 US-08-232-463-14	Sequence 14, Appl
2	57.6	3.0	3666	2 US-08-682-517-13	Sequence 13, Appl
3	57.6	3.0	3666	2 US-08-682-517-14	Sequence 14, Appl
4	57.6	3.0	4197	2 US-08-682-517-7	Sequence 7, Appl
5	57.6	3.0	4197	2 US-08-682-517-8	Sequence 8, Appl
6	49.6	2.6	2793	1 US-08-209-747-1	Sequence 1, Appl
7	49.6	2.6	2793	1 US-08-458-298-1	Sequence 1, Appl
8	46.8	2.4	5361	4 US-08-973-462-2	Sequence 2, Appl
9	46.8	2.4	6152	4 US-08-973-462-1	Sequence 1, Appl
C 10	46.8	2.4	543	6 5273901-6	Patent No. 5273901
C 11	45.8	2.4	397	3 US-09-253-691-3	Sequence 3, Appl
12	45.8	2.4	5319	1 US-08-169-927-1	Sequence 1, Appl
13	43.6	2.3	1891	4 US-08-973-462-3	Sequence 3, Appl
14	43.4	2.3	789	3 US-08-483-857-3	Sequence 3, Appl
15	43.4	2.3	1800	3 US-08-483-857-1	Sequence 1, Appl
16	43.2	2.3	1173	4 US-09-134-001C-281	Sequence 281, App
17	43.2	2.3	2169	4 US-09-434-408-3	Sequence 3, Appl
C 18	42.6	2.2	234	1 US-08-469-802B-3	Sequence 3, Appl
C 19	42.6	2.2	234	2 US-08-267-803B-3	Sequence 3, Appl
C 20	41.6	2.2	203	4 US-09-043-303-7	Sequence 7, Appl
C 21	41.6	2.2	477	4 US-09-135-994-1	Sequence 1, Appl
C 22	41.6	2.2	3168	4 US-09-165-239A-3	Sequence 3, Appl
C 23	41.4	2.2	740	4 US-09-451-117-1	Sequence 1, Appl
24	41.4	2.2	2472	4 US-09-134-001C-1244	Sequence 1244, App
C 25	40.8	2.1	1267	4 US-09-071-035-483	Sequence 483, App
C 26	40.8	2.1	1356	4 US-09-071-035-481	Sequence 481, App
27	40.4	2.1	2830	2 US-09-010-928B-1	Sequence 1, Appl

C 28	40.4	2.1	8585	1 US-08-030-096-3	Sequence 3, Appl
29	39.4	2.1	10851	2 US-08-286-819A-16	Sequence 16, Appl
30	39.4	2.1	10851	3 US-08-980-357-16	Sequence 16, Appl
31	39.2	2.0	789	1 US-08-436-748-4	Sequence 4, Appl
32	39.2	2.0	1800	1 US-08-436-748-1	Sequence 1, Appl
C 33	39	2.0	533	6 5482709-5	Patent No. 5482709
C 34	39	2.0	4248	3 US-08-678-614-1	Sequence 1, Appl
35	39	2.0	4287	1 US-08-038-682-7	Sequence 7, Appl
36	39	2.0	4287	1 US-08-302-832-7	Sequence 7, Appl
37	39	2.0	4287	2 US-08-530-198-7	Sequence 7, Appl
38	39	2.0	4287	2 US-08-469-880-7	Sequence 7, Appl
39	39	2.0	4287	2 US-08-728-470-7	Sequence 7, Appl
40	39	2.0	4287	4 US-08-719-641-7	Sequence 7, Appl
41	39	2.0	4702	1 US-08-038-682-8	Sequence 8, Appl
42	39	2.0	4702	1 US-08-302-832-8	Sequence 8, Appl
43	39	2.0	4702	2 US-08-530-198-8	Sequence 8, Appl
44	39	2.0	4702	2 US-08-469-880-8	Sequence 8, Appl
45	39	2.0	4702	2 US-08-728-470-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match

3.1%; Score 60.4; DB 1; Length 7218;

QY	296	TAACTGGTGATATTTCATTAGGTTCCAGTGTAGATACGGGAGGAGCTAATAAACCTTGCAG	355
Db	2909	TAACTGCTGCTACATTAAACAGTTACTTATGCGAGTGTAAAAATGCTGCAGGTGTGCTG	2968
QY	356	TTAATATTGATGATGTTTTAACTTTAACCTTTAACAGGTACC GGTA CTG CAGCCTACGGTG	415
Db	2969	AAAATTTACTGCTAGCGTAACAATTAAAGAAAC---TACTGGAGCAATTACTTCTGATA	3035
QY	416	CAAATCCTGCGTTGTATTCCAAAGSTGGACAAGCTGCTGAATAATACATATATCTGCT	475
Db	3026	CATTTACACAAGGTGTTATACCATCAGCAGCTACAGCAGCTCAATATACTTCTTAATCAA	3085
QY	476	TAGTAATATAACTCTAGGTGGAGCGAATGCCGGTTTGACTATTGCTTCAGATCCAGATG	535
Db	3086	TTGCTGCAGATTATACATTTTGCACAGGTGAAGGATTCATTTTAATATTGATAATGCTG	3145
QY	536	TATTAGGACCAATAAGCGTTTGCAGGAAATATATAGATGGAGGAGGTATAATAACTGACAATA	595
Db	3146	GTGCTCAAGTAATTAATCTTAGCAGSTAAAAAGGTGCACAAGGT---GTAGCTGATGCTA	3202
QY	596	CAGATGCTGCCATTAAACGGGAACAATAAGGTAATACTAATCCGCGACGCTCAAAATAAGCAATTG	655
Db	3203	TCAATGCTACATTTGCGAGGTACTGCAACTGTTTCTGGAGACAAAGTAGTTATTAAATCAG	3262
QY	656	GAGCAAGTAGCGCTTCTCTTTGGAGGGCGAGTTATTAAAGCCCACTACGACTAAGTTAAACAA	715
Db	3263	CTACAACAGGTGTGTGTTCTGGAAGTTGAAGTTACATCTCTTCTGTTAATCAAGTATTAA	3322
QY	716	ATGCAG	721
Db	3323	ATGCAG	3328

; Sequence 13, Application US/08682517

```

RESULT 3
US-08-682-517-14
; Sequence 14, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3666
US-08-682-517-14

```

Query Match	3.0%;	Score 57.6;	DB 2;	Length 3666;
Best Local Similarity	47.0%;	Pred. No. 1.8e-05;		
Matches 285;	Conservative	0;	Mismatches 309;	Indels 12; Gaps

QY	116	CTGCAAGACCGTAACTGCTGTATGTCGAGAGCTTGCAGCCGGACAAATATAGTCCGT	175
Db	2735	CTTTAGGAATTTCAATAGCTGATGCAGATCTTAATGTAAAGTGCACCAACTCTTGATACTG	2794
QY	176	GAGCCCGGTGCTTTTGTAGCGGGTTCTACTTTACAATATACCGGTGCTTTTACGGTTACTG	235

Db 2795 CAACGTGTTTCATTAAG-----ATAGTGAATAATTCATATCTCTTACATAGTTG 2848
 Qy 236 ATGCTGAGCTAGTCTTCGTCATTAGATTTAAATAATTTTCACAGGTCTTTTTCAG 295
 Db 2849 AAACGTGGTCTAATACAGGTGATTTGCTACAACTGTTCAAGCTGGTCAATATCTT 2908
 Qy 296 TAACTGGTGATATTCATAGTTTCAGTGGTAGATACAGGGAGAGCTAATAAATTCGAG 355
 Db 2909 TAACTGGTGATATTCATAGTTTCAGTGGTAGATACAGGGAGAGCTAATAAATTCGAG 2968
 Qy 356 TTAATATTGATGATGTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTT 415
 Db 2969 AAAATATTACTGCTAGCGTAACATTAAGAAAC---TACTGGAGCAATTAATCTTGATA 3025
 Qy 416 CAAATCCCTGCGTGTATTCCAAAGTGGACAAAGCTGCTGCTAATAATACATATACATGCTT 475
 Db 3026 CATTTACACAAGGTGATTAACCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 3085
 Qy 476 TAGGTAAATATACTCTAGTGGAGGGAATGCCGGTTTGACTTATTTGCTTCAGATCCAGATG 535
 Db 3086 TTGCTGCAGATTATACATTTGCAACAGGTGAAGGATTCACITTTAAATATTGATAATGCTG 3145
 Qy 536 TATTAGGACCAATACGCTTGAGGAAATATAGATGGAGGAGGTATATAAATGACAATA 595
 Db 3146 GTGCTCAAGTAATTAACCTTAGCAGGTAAAGGAGTGCACAAGGT---GTAGCTGATGCTA 3202
 Qy 596 CAGATGCTGCCATTAACGGAACAATAGGTAACTAATCCGCGAGCTCAAAATAGCAATG 655
 Db 3203 TCAATGCTACATTTGCGAGTACTGCACTGTTCTTGGAGACAAGTAGTTTAAATCAG 3262
 Qy 656 GAGCAAGTACGCTTCTTCTTGGAGGGGCAAGTTATTAAAGCCACTACGACTAAAGTTAAACAA 715
 Db 3263 CTACAACAGGTGCTGTTCTGGAAGTTGAAGTTACATTTCTTCTGTTAATCAAGTATTAA 3322
 Qy 716 ATGCAG 721
 Db 3323 ATGCAG 3328

RESULT 4
 US-08-682-517-7
 ; Sequence 7, Application US/08682517
 ; Patent No. 5874267
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Expression of surface layer proteins
 ; NUMBER OF SEQUENCES: 25
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/682,517
 ; FILING DATE:
 ; CLASSIFICATION:
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4197 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-682-517-7
 Query Match 3.0%; Score 57.6; DB 2; Length 4197;
 Best Local Similarity 47.0%; Pred. No. 1.9e-05;
 Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;
 Qy 116 CTGCAAGAACCGTAACCTGCTGATGGTGGAGAGCTTCAGCGGCAACAATATAGGTCCTG 175
 Db 2919 CTTTGAATATTCATTAGCTGATGCAATCTTAATGAAGTCAACAATCTGTTGATAGT 2978

Qy 176 GAGCGGTGCTTTTGTAGCGGTTCTACTTACAAATATACCGGTGCTTTTACGGTTACTG 235
 Db 2979 CAACGTGTTTCATTAAG-----ATAGTGAATAATTCATATCTCTTACATAGTTG 3032
 Qy 236 ATGCTGAGCTAGTCTTCGTCATTAGATTTAAATAATTTTCACAGGTCTTTTTCAG 295
 Db 3033 AAACGTGGTCTAATACAGGTGATTTGCTACAACTGTTCAAGCTGGTCAATATCTTCTT 3092
 Qy 296 TAACTGGTGATATTCATAGTTTCAGTGGTAGATACAGGGAGAGCTAATAAATTCGAG 355
 Db 3093 TAACTGGTGATATTCATAGTTTCAGTGGTAGATACAGGGAGAGCTAATAAATTCGAG 3152
 Qy 356 TTAATATTGATGATGTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTT 415
 Db 3153 AAAATATTACTGCTAGCGTAACATTAAGAAAC---TACTGGAGCAATTAATCTTGATA 3209
 Qy 416 CAAATCCCTGCGTGTATTCCAAAGTGGACAAAGCTGCTGCTAATAATACATATACATGCTT 475
 Db 3210 CATTTACACAAGGTGATTAACCTTAGCAGGTAAAGGAGTGCACAAGGT---GTAGCTGATGCTA 3269
 Qy 476 TAGGTAAATATACTCTAGTGGAGGGAATGCCGGTTTGACTTATTTGCTTCAGATCCAGATG 535
 Db 3270 TTGCTGCAGATTATACATTTGCAACAGGTGAAGGATTCACITTTAAATATTGATAATGCTG 3329
 Qy 536 TATTAGGACCAATACGCTTGAGGAAATATAGATGGAGGAGGTATATAAATGACAATA 595
 Db 3330 GTGCTCAAGTAATTAACCTTAGCAGGTAAAGGAGTGCACAAGGT---GTAGCTGATGCTA 3386
 Qy 596 CAGATGCTGCCATTAACGGAACAATAGGTAACTAATCCGCGAGCTCAAAATAGCAATG 655
 Db 3387 TCAATGCTACATTTGCGAGTACTGCACTGTTCTTGGAGACAAGTAGTTTAAATCAG 3446
 Qy 656 GAGCAAGTACGCTTCTTCTTGGAGGGGCAAGTTATTAAAGCCACTACGACTAAAGTTAAACAA 715
 Db 3447 CTACAACAGGTGCTGTTCTGGAAGTTGAAGTTACATTTCTTCTGTTAATCAAGTATTAA 3506
 Qy 716 ATGCAG 721
 Db 3507 ATGCAG 3512

RESULT 5
 US-08-682-517-8
 ; Sequence 8, Application US/08682517
 ; Patent No. 5874267
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Expression of surface layer proteins
 ; NUMBER OF SEQUENCES: 25
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/682,517
 ; FILING DATE:
 ; CLASSIFICATION:
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4197 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ORIGINAL SOURCE:
 ; ORGANISM: Bacillus sphaericus
 ; INDIVIDUAL ISOLATE: P-1
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 95..3850
 ; FEATURE:
 ; NAME/KEY: mat_peptide

```

; LOCATION: 185..3850
; FEATURE:
; NAME/KEY: sig_pepti
; LOCATION: 95..184
US-08-682-517-8

```

Query Match	3.0%; Score 57.6; DB 2; Length 4197;
Best Local Similarity	47.0%; Pred. No. 1.9e-05;
Matches	285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;
QY 116	CTGCAAGAACCGTAACCTGCTGATCGTCAGAGCTTCGAGCCGGACCAAAATATAGTGCCTG 175
DB	
QY 2919	CTTTAGAAATTCATTAAGCTGATGCAGATCTTAATGTAAGTGCAACAACCTGTTGATAGT 2978
DB	
QY 176	GAGCCGGTGCTTTTGTAGCCGGTTCTACTTTTACAATATACCGGTGCTTTTACGGTTACTG 235
DB	
QY 2979	CAACTGTTTCATTAAG-----ATAGTGCAAATAATTCATTTATCTTTACATTAGTTG 3032
DB	
QY 236	ATGCTGACGTAAGTTCGTGCATTAGATTTAAATAATTTTGCAGAGGCTCTTTTTCAG 295
DB	
QY 3033	AAACTGGTGCTAATACAGGTGTTATTTCTACACTGTTCAGCTGGTACATTTATCTTCTT 3092
DB	
QY 296	TAACTGCTGATATTTCATTAGGTTTCAGTGTAGATACGGGAGGAGCTAAATAAACTTCGAC 355
DB	
QY 3093	TAACTGCTGTACATTAACAGTTACTTTATGCAGATGCTAAAATGCTGCAGGTGTTGCTG 3152
DB	
QY 356	TTAATATTGATGTTTAACTTTAACTTTTACAGTACGGTACTGCGACCCCTACCGTG 415
DB	
QY 3153	AAATAATTACTGTAGCGGTAACTTAAAGAAAAC---TACTGGAGCAATTACTTCTGATA 3209
DB	
QY 416	CAAACTCTGCGTTGTTATTTCCAAAGTGCACAAGCTGCTGCTTAATAATACATATACTGCTT 475
DB	
QY 3210	CATTTACACAAGGTGATTACCATCAGCAGCTACAGCAGCTGATATACTTCTAAATCAA 3269
DB	
QY 476	TAGTAAATATACTTAGTGGAGCGAATCCCGGTTTGACTATTCGTTTCAGATCCAGATG 535
DB	
QY 3270	TTGCTGCAGATTATACATTTTGCACAGGTGAAGGATTCATTTAAATATTGTAANTGCTG 3329
DB	
QY 536	TATTAGGACCAATAAACCTTTCAGGAAATATAGATGAGGAGGTATTAATAACTGCACAATA 595
DB	
QY 3330	GTGCTCAAGTAATTAACTTAGCAGGTAAAAAGGTGCACAAGGT--GTAGCTGATGCTA 3386
DB	
QY 596	CAGATGCTGCCATTACGGGAACAATAGGTAATACTAATCCGGCAGCTCAAAATGAAGATTG 655
DB	
QY 3387	TCAATGCTACATTTTGCAGGTACTGCAACTGTTTCTGGAGACAAGTAGTTATTAAATCAG 3446
DB	
QY 656	GAGCAAGTACGCTTCTCTTGGAGGGCAGTTTATTAAGCCACTACGACTAAGTTAAACA 715
DB	
QY 3447	CTACACAGGTGTGGTCTGAAGTTGAAGTTACATCTCTTCTGTTAATCAAGATTAA 3506
DB	
QY 716	ATGCAG 721
DB	
QY 3507	ATGCAG 3512
DB	

```

RESULT 6
US-08-209-747-1
: Sequence 1, Application US/08209747
: Patent No. 5733771
: GENERAL INFORMATION:
: APPLICANT: Lewis, Randolph V.
: APPLICANT: Colgin, Mark
: TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
: TITLE OF INVENTION: Silk Proteins
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch
: STREET: P.O. Box 747
: City: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22040-3487
: COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/209,747
FILING DATE: 14-MAR-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-104P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2793 base pairs
TYPE: nucleic acid
STRADEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Nephila clavipes
TISSUE TYPE: minor ampullate gland
FEATURE:
NAME/KEY: CDS
LOCATION: 183..2675
OTHER INFORMATION: /product="N. clavipes minor"
OTHER INFORMATION: ampullate silk protein"
US-08-209-747-1

Query Match	2.6%;	Score 49.6;	DB 1;	Length 2793;
Best Local Similarity	41.5%;	Pred. No. 0.0019;		
Matches 157;	Conservative 0;	Mismatches 221;	Indels 0;	Gaps

Qy	95	CCGGTAGTCGAGTCCCTTGTTGTCTGCACAAGACCGTAACCTGCTCATGCTCAGAGCTTGCGAG	154
Db	892	CTGTGCTGGAGCTGGAGCCGCTGCAGGTGCCAGGACGAGGACTGGAAGCTACGGTGTGTC.	951
Qy	155	CCGGAACAAATATAGAGTCCTGGAGCCGGTGCTTTTGTAGCGGGTCTACTTTTACAATA	214
Db	952	AAGGATACGCTGCCGAGCAGGAGCTGGTCTGCTGCAGCTGCANNNNNNNNNNNNNN	1011
Qy	215	CCGGTGCCTTTTACGGTTACTGATGCTGAGCTAAGCTTCGTGCTATTAGATTAAATAAT	274
Db	1012	NNNNNNNNNNNNNNNNNNNNNGTGCAGGTGCAGGTGCTGGATATGGTGGACAG	1071
Qy	275	TTCGAGCAGGCTCTTTTTTCAGTAACCTGGTGATATTTCAATTAGGTTCACTGGTAGATACGG	334
Db	1072	CGCGATATGTCGCCGAGCAGGAGCTGGTGGCGCTGCTGCTGGTGCAGGAGCTGGAG	1131
Qy	335	GAGGACCTAATAACCTTGCAGTTTAATATATGATGATGTTTAACTTAACTTTACAGGTA	394
Db	1132	GTCTCGTGGTTACCGCTAGAGGTGCTGGTCTGGAGCTGGAGCCGCTCAGGTGCAGGAG	1191
Qy	395	CCGGTACTGCAGCTACCGTGCAAAATCCCTGCCCTTCTTATCCAGGTCGACAACTGCTG	454
Db	1192	CAGGACTGGAGGCTACGGTGGTGCAAGTGGATACGGTGCCGAGCAGGAGCTGCTGCAG	1251
Qy	455	CTAATAATACATATAGT	472
Db	1252	CTGCTGGAGCAGGAGCTG	1269

RESULT 7

US-08-458-298-1

; Sequence 1, Application US/08458298

; Patent No. 5756677

; GENERAL INFORMATION:

; APPLICANT: Lewis, Randolph V.

; APPLICANT: Colgin, Mark

;; TITLE OF INVENTION: CDNAS Encoding Minor Ampullate Spider

;; TITLE OF INVENTION: Silk Proteins

;; NUMBER OF SEQUENCES: 56

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Birch, Stewart, Kolasch & Birch

;; STREET: P.O. Box 747

;; CITY: Falls Church

;; STATE: Virginia

;; COUNTRY: USA

;; ZIP: 22040-3487

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA: US/08/458,298

;; FILING DATE: 02-JUN-1995

;; CLASSIFICATION: 530

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/209,747

;; FILING DATE: 14-MAR-1994

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Murphy Jr., Gerald M.

;; REGISTRATION NUMBER: 28,977

;; REFERENCE/DOCKET NUMBER: 1447-104P

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 703-205-8000

;; TELEFAX: 703-205-8050

;; INFORMATION FOR SEQ ID NO: 1:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 2793 base pairs

;; TYPE: nucleic acid

;; STRANDEDNESS: double

;; TOPOLOGY: linear

;; MOLECULE TYPE: cDNA

;; HYPOTHETICAL: NO

;; ORIGINAL SOURCE:

;; ORGANISM: Nephila clavipes

;; TISSUE TYPE: minor ampullate gland

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: 183..2675

;; OTHER INFORMATION: /product="N. clavipes minor

;; OTHER INFORMATION: ampullate silk protein"

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

Db 1192 CAGGAGCTGGAGCTACGGTGGTCAAAAGTGGATACGGTCCGAGCAGGAGCTGCTGCAG 1251

Qy 455 CTAATAATACATATACG 472

Db 1252 CTGCTGGAGCAGGAGCTG 1269

RESULT 8

US-08-973-462-2

; Sequence 2, Application US/08973462B

; Patent NO. 6191270

; GENERAL INFORMATION:

; APPLICANT: DRUILHE, PIERRE

; APPLICANT: DAUBERSIES, PIERRE

; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES

; FILE REFERENCE: 0660-0125-0 PCT

; CURRENT APPLICATION NUMBER: US/08/973,462B

; CURRENT FILING DATE: 1998-02-06

; EARLIER APPLICATION NUMBER: PCT/FR96/00894

; EARLIER FILING DATE: 1996-06-12

; EARLIER APPLICATION NUMBER: FR 95/07007

; EARLIER FILING DATE: 1995-06-13

; NUMBER OF SEQ ID NOS: 29

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 5361

; TYPE: DNA

; ORGANISM: P. falciparum

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(5361)

US-08-973-462-2

Query Match 2.4%; Score 46.8; DB 4; Length 5361;

Best Local Similarity 40.9%; Pred. No. 0.013; Indels 9; Gaps 2;

Matches 527; Conservative 0; Mismatches 752;

Qy 75 AACCCGACGGATAATGCTAAACCGGTAGTGGAGTCTCTGGTGTGCAAGAACCGTAACCTGC 134

Db 855 AATCGTAGCTCCAAGTGTGTAGAAAGTGTGGCTCCAAGTGTGAAGAAAGTGTAGAAGA 914

Qy 135 TGATGGTCAGAGCTTCAGCGCGGAAACAATATAGTCTCTGGAGCGGTGCTTTTGTAGC 194

Db 915 AAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGTGAATGTTGAAGA 974

Qy 195 GGGTCTTACTTTTACAATATACCGGTGCTTTACGGTGTACTGATGCTGACGTAAGTGTTCG 254

Db 975 AAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAATGCTAGCTCC 1034

Qy 255 TGCATTAGATTTAAATAATTTTGCAGCAGGTCTTTTTCAGTAAGTGTGATGATTTTCAAT 314

Db 1035 AACTGTTGAAGAAA-----TCGTAGCTCCAAGTGTGAAGAAATGTTAGCTCCAAGTGT 1088

Qy 315 AGGTTTCAGTGTAGATACGGGAGGAGCTAATAAAGTGTGCAAGTGTGATGATGATGTTT 374

Db 1089 TGTAGAAAGTGTGGCTCCAAGTGTGAAGAAAGTGTGAAGAAAGTGTGAAGAAAGTGT 1148

Qy 375 AACCTTAAGTTTAAACAGGTACCGGTACTGACGCTACGGTACGCTGCAATCTCTGCTGTTAT 434

Db 1149 AGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTT 1208

Qy 435 CCAAGGTGACAAAGTGTGCTGCTTAATACATATATCTGCTTTAGTGAATATATACCTCTAGG 494

Db 1209 TGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGT 1268

Qy 495 TGGACGGAATCCGGTTTTCAGTATTTGCTTTCAGATCCAGATGATTTAGGACCAATACCGT 554

Db 1269 AGCTCCAAGTGTGAAGAAATCGTAGCTCCAAGTGTGAAGAAATGTTAGCTCCAAGTGT 1328

Qy 555 TCCAGAAATATAGATGGAGGAGGTATATAACTGACATACAGATGCTGCCATTAACCG 614

Db 1329 TGTAGAAAGTGTGGCTCCAAGTGTGAAGAAAGTGTGAAGAAAGTGTGAAGAAAGTGT 1388

```

QY 615 AACATAGGTAACTAATCCGGCAGCTCAAAATAGCAATGGAGCAAGTACGCTTTCTCT 674
Db 1389 AGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGT 1448
QY 675 TGGAGGGCAGTTATTAAGCCACTACGACTAAGTTACAAATGACGCGCGGTATTAC 734
Db 1449 TGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGT 1508
QY 735 ACTTACAAATGCAATGAGTATTAACAGGTGCTGTTGATACACACACAGCGGTGATGA 794
Db 1509 AGCTGAAATGTTGAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGT 1568
QY 795 TGTAGGTGCTTAAATTTAAACGGAGCGTTGAGTCAAGTAACTGGAAATATAGTTAATAC 854
Db 1569 TGAAGAAATGCTAGCTCCAACTGTTGAAGAAATGTTGAGTCCAACTGTTGAGAAAGTGT 1628
QY 855 AAATTCATTAGCGCAATTAAGTGTAGGAGCAGGTACGCTAGCTTAGGGGAGCGGTAT 914
Db 1629 GGCTCCAAAGTGTGAAGAAAGTGTGAAGAAAGTGTGAAGAAAGTGTAGCTGAAATGT 1688
QY 915 TAAAGCCACTACGACTAAGTTAAACGAATGCAGCGTCTGTTAATACACTTACAAATGCAGT 974
Db 1689 TGAAGAAAGTGTAGCTGA---AAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAG 1745
QY 975 ATTAACAGGTGCTGTTGATACACACACAGCGGTGATTAATGTAGGTGCTGTAATCTAAG 1034
Db 1746 TGTAGCTGAAATGTTGAAGAAAGTGTAGCTCCAACTGTTGAAGAAAGTGTAGCTCCAACT 1805
QY 1035 TGGAGCAATGAGTCAAGTAAACCGGCAATATAGTAAATACAAATTCATTAGCCACGATAA 1094
Db 1806 TGTGAAGAAATGTTAGCTCCAAAGTGTGAGAAAGTGTGCTCCAACTGTTGAAGAAAG 1865
QY 1095 TATAGGACAGGTGCTGCTTGGATGGAGCGGTTATTAAAGCTACTAGCACTAAGTT 1154
Db 1866 TGTGAAGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAAT 1925
QY 1155 AACAGATGATCGTCAGTATTAAATTTACGAATCCTGTAGTAGTACCGGAGCAATAGA 1214
Db 1926 TGTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTCCAACTGTTGAAGAAAT 1985
QY 1215 TAATACCGGTAAATCCAAATAGGTGTTGTAATCTTTACCGGAGCAAGTACGGTAAACCGA 1274
Db 1986 CGTAGCTCCAACTGTTGAAGAAATGTTAGCTCCAAAGTGTGTAAGAAAGTGTGCTCCAA 2045
QY 1275 TAATATAGGTAAACCGCAGTATTAGCAGAGGTAAAGCGTAGGACGAGGTGCTGCAAAAT 1334
Db 2046 TGTGAAGAAAGTGTAGCAAGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAG 2105
QY 1335 ACAAGCGGAGTGTAAAGCGCAATGCA 1362
Db 2106 TGTAGCTGAAATGTTGAAGAAAGTGTGA 2133

```

```

RESULT 9
US-08-973-462-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUIHE, PIERRE
; APPLICANT: DAUBESIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-BRYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA

```

ORGANISM: P. falciparum
US-08-973-462-1

Query Match 2.4%; Score 46.8; DB 4; Length 6152;
Best Local Similarity 40.9%; Pred. No. 0.013;
Matches 527; Conservative 0; Mismatches 752; Indels 9; Gaps 2;

```

QY 75 AACCGCAGCGATTAATGCTAAACCGGTAGTGGAGTCTTGGTCTGCTCAAGAACCGTAACGTC 134
Db 1099 AATCGTAGCTCCAAGTGTGTAGAAAGTGTGGCTCCAAGTGTGAAGAAAGTGTGAAGA 1158
QY 135 TCATGGTCAGAGCTTGCAGCGGCAACAAATATAGTCTCTCGAGCGGTGCTTTTGTAGC 194
Db 1159 AAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGA 1218
QY 195 GGGTTCACCTTTACAAATATACCGGTGCTTTTACGCTTACTGATGCTGACGTAAGTGTTCG 254
Db 1219 AAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTCC 1278
QY 255 TGCATTAGATTTAATAATTTTGCAGCAGGTCTTTTTCAGTAACTGCTGATATTTCATT 314
Db 1279 AACTGTTGAAGAAA-----TCGTAGCTCCAAGTGTGTGAAGAAATGTTAGCTCCAAGTGT 1332
QY 315 AGTTCAGTGTAGTATACGGGAGGAGCTTAATAAATTTGCAGTAAATATTGATGATGTTT 374
Db 1333 TGTAGAAAGTGTGCTCCAAGTGTGAAGAAAGTGTGAAGAAAGTGTGAAGAAAGTGT 1392
QY 375 AACCTTAACCTTACAGGTACCGGTACTGCGAGCTAGCGTCAAAATCCTGGGTTGTTATT 434
Db 1393 AGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGT 1452
QY 435 CCAAGTGGCAAGCTGCTGCTAATAATATACATATACCTTTAGTGTAAATATAAATCTAGG 494
Db 1453 TGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGT 1512
QY 495 TGGAGCAATGCGGTTTGGCTTTCAGATCCAGATGTTATTAGGACCAATTAACGCT 554
Db 1513 AGCTCCAACTGTTGAAGAAAGTGTAGCTCCAAGTGTGAAGAAAGTGTAGCTCCAAGTGT 1572
QY 555 TGCAGGAAATATAGATGAGGAGGTATTAATACATACAGTGTGAAGAAAGTGTGAAGAAAGTGT 614
Db 1573 TGTAGAAAGTGTGCTCCAAGTGTGAAGAAAGTGTGAAGAAAGTGTGAAGAAAGTGT 1632
QY 615 AACAAATAGTAACTAATCCGCGAGCTCAAATAAGCAATTTGGACAAAGTACGCTTTCTCT 674
Db 1633 AGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGT 1692
QY 675 TGGAGGGCAGTTATTAAAGCCACTACGACTAAGTTAAACAAATGCGCGCGGTATTAAAC 734
Db 1693 TGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGT 1752
QY 735 ACTTACAAATGCAATGCAGTATTAAACAGGTGCTGTTGATAAACACACACGCGGTGATGA 794
Db 1753 AGCTGAAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAGTGTAGCTCCAAGTGT 1812
QY 795 TGTAGGTGCTTAAATTTAAACGGAGCGGTGTAGTCAAGTAACTGGAAATATAGGTAATAC 854
Db 1813 TGAAGAAATGCTAGCTCCAACTGTTGAAGAAATGTTAGCTCCAAGTGTGTAGAAAGTGT 1872
QY 855 AAATTCATTAGCGCAATTAAGTGTAGGAGCAGGTACGCTACGCTTAGGGGAGCGGTAT 914
Db 1873 GGCTCCAAAGTGTGAAGAAAGTGTGAAGAAAGTGTGAAGAAAGTGTAGCTGAAATGT 1932
QY 915 TAAAGCCACTACGACTAAGTTAAACGAATGCAGGTGCTGTTAATACACTTACAAATGCAGT 974
Db 1933 TGAAGAAAGTGTAGCTGA---AAATGTTGAAGAAAGTGTAGCTGAAATGTTGAAGAAAG 1989
QY 975 ATTAACAGGTGCTGTTGATACACACACGCGGTGATTAATGTAGGTGCTGCTAATCTAAG 1034
Db 1990 TGTAGCTGAAATGTTGAAGAAAGTGTAGCTCCAAGTGTGAAGAAAGTGTAGCTCCAAC 2049
QY 1035 TGGAGCAATGAGTCAAGTAAACCGGCAATATAGTAAATACAAATTCATTAGCCACGATAA 1094

```

Db 2050 TGTTCGAAGAAATGTAGCTCCAAAGTGTGTGAGAAAGTGTGGCTCCCAAGTGTGGAAGAAG 2109
QY 1095 TATAGAGCAGGTGGCTACCTTGGATGGAGCGGTTATTAAAGCTACTACGACTAAGTT 1154
Db 2110 TGTAGAAGAAATGTGTGAAGAAAGTGTAGCTGAAATGTGTGAAGAAAGTGTAGCTGAA 2169
QY 1155 AACAGATGATCGTGCAGTATTATATTTACGAATCCTGTAGTAGTACCGGGAGCAATAGA 1214
Db 2170 TGTGAAGAAAGTGTAGCTGAAATGTGTGAAGAAATCGTAGCTCCAACTGTTGAAGAAT 2229
QY 1215 TAATACCGGTAAATGCAATAAGAGTGTGTATCTTTTACCGGAGCAAGTACGGTAACCGA 1274
Db 2230 CGTAGCTCAACTGTTGAAGAAATGTAGCTCCAAAGTGTGTGAAGAAAGTGTGGCTCCAAG 2289
QY 1275 TAATATAGGTACACGGCAGTATTAGCAGAGTACGAGTACGAGTGTGGCTGCTCAAT 1334
Db 2290 TGTGAAGAAAGTGTAGAGAAATGTGTGAAGAAAGTGTAGCTGAAATGTGTGAAGAAG 2349
QY 1335 ACAAGCGGAGTAGTAAAGCGAATGCA 1362
Db 2350 TGTAGCTGAAATGTGTGAAGAAAGTGA 2377

RESULT 10
5273901-6/c
; Patent No. 5273901
; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILLSON,
; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
; SPOOROZITE 21.5 KB ANTIGEN, AC-6B
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/581,693
; FILING DATE: 12-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1988
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO: 6:
; LENGTH: 543
5273901-6

Query Match 2.4%; Score 46.4; DB 6; Length 543;
Best Local Similarity 53.3%; Pred. No. 0.0067;
Matches 98; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 110 TTGGTCTGCAAGAACCGTAACGTGATGTTGCGAGCTTCAGCGCGCAACAAATATAG 169
Db 244 TTGCTCTGCTGCTGCGAGCTGTGGCTGCGAGCTGGTGGCTGAGCTGAGCTGGTG 185
QY 170 GTCTGAGCGCGTGTGTTTGTAGCGGGTTCTACTTTTACAATATACCGTGTCTTTACGG 229
Db 184 TTGCTAGCTGCTGCTGTGCGAGCTGCTTCTCTCTGCTGCTGCTGCGAGTGGTG 125
QY 230 TTAGTATGCTGACGTAAGTGTTCGGCAATATAGATTTAAATATTTTTCAGCAGGTCTTT 289
Db 124 CTGCTGTGCTGCTGCGAGTGTGCGGCTCTTTCTTCTTCTTCTTCTTCTTCTTCTC 65
QY 290 TTTC 293
Db 64 TCTC 61

RESULT 11
US-09-253-691-3/c
; Sequence 3, Application US/09253691
; Patent No. 6124100
; GENERAL INFORMATION:
; APPLICANT: Dong Kyu JIN

; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; TITLE OF INVENTION: Using Trinucleotide Repeats Sequence
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691
; EARLIER FILING DATE: 1999-02-22
; EARLIER APPLICATION NUMBER: KR 98-6,278
; EARLIER FILING DATE: 1996-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 3
; LENGTH: 397
; TYPE: DNA
; ORGANISM: human
US-09-253-691-3

Query Match 2.4%; Score 45.8; DB 3; Length 397;
Best Local Similarity 48.0%; Pred. No. 0.0085;
Matches 131; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
QY 76 ACCGAGCGATAATGCTAACCGGTAGTGGAGTCCCTTGGTGTGCAAGAACCGTAACGTCT 135
Db 377 AACTCTGCTGCTAGTAGGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 318
QY 136 GATGTCAGAGCTTTCAGCGGAAACAAATATAGTCTTCGAGCCCGGCTCTTTTGTAGCG 195
Db 317 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 258
QY 196 GGTTCCTACTTTACATATACCGGTGCTTTTACGGTTACTGATGCTGACGTAAGTCTCT 255
Db 257 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 198
QY 256 GCATTAGATTTAAATAATTTTCAGCAGGTCTTTTTCAGTAACCTGCTGATATTTCAATTA 315
Db 197 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 138
QY 316 GGTTCAGTGTAGATACGGGAGGAGCTAATAAA 348
Db 137 GTTGCTGCTTTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 105

RESULT 12
US-08-169-927-1
; Sequence 1, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dobson, Michael E.
; APPLICANT: Ching, Wei Mei
; APPLICANT: Dasch, Gregory A
; TITLE OF INVENTION: Gene and Protein Applicable to the
; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Counsel, Naval Medical R & D Command
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,927
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/742,128
; FILING DATE: 08/09/91
; ATTORNEY/AGENT INFORMATION:

```

: NAME: Spavack, A. David
: REGISTRATION NUMBER: 24,743
: REFERENCE/DOCKET NUMBER: 75,976
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 295-6759
: TELEFAX: (301) 295-1022
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5319 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: circular
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: internal
: ORIGINAL SOURCE:
: ORGANISM: Rickettsia prowazekii
: STRAIN: Breinl
: FEATURE:
: NAME/KEY: -35_signal
: LOCATION: 340..345
: FEATURE:
: NAME/KEY: -10_signal
: LOCATION: 363..368
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 391..5226
: FEATURE:
: NAME/KEY: RBS
: LOCATION: 379..386
: FEATURE:
: NAME/KEY: stem_loop
: LOCATION: 5270..5306
: PUBLICATION INFORMATION:
: AUTHORS: Carl, M.
: AUTHORS: Dobson, M. E.
: AUTHORS: Ching, W. M.
: AUTHORS: Dasch, G. A.
: TITLE: Characterization of the
: TITLE: protective S-layer prote
: TITLE: prowazekii; presence of a
: TITLE: homolog in rickettsia ty
: JOURNAL: Proc. Natl. Acad. Sci.
: DATE: 1990
: RELEVANT RESIDUES IN SEQ ID NO:
: US-08-1669-927-1

```

Query Match	2.4%	Score 45.6;	DB 1;	Length 5319;
Best Local Similarity	49.6%;	Pred. No. 0.026;		
Matches 117;	Conservative 0;	Mismatches 119;	Indels 0;	Gaps 0;
QY 1488	AGGAAGCCTAGCTCGCAATAATATAGATTTTGGAGCTGCCAGTAATTTAGAGTTTAAACGG	1547		
Db				
QY 1917	AGCTAAGATGCTTCAAAAATATTAGCACTCGATGGCGCAATATTATCGGGGCTAATGT	1976		
Db				
QY 1548	TCCTCGCGGTAGAATTTATTAACTTAATCGGAACATATAGCAAAACGGTAATAATGCTACACT	1607		
Db				
QY 1977	TGTTGTTGCAATTCATTTTCAAGCTAAACGGTGGTACTATTAAATTAACAATACTCAAAA	2036		
Db				
QY 1608	TAATATTAATGCTGCTGTCACAGTGATTGCAAAATGATGTTAGTAGTAGGCACAGTTGCACA	1667		
Db				
QY 2037	TAATATGTAGTTAAATTTTGCATTTAGATATAACTACTGATAAACAAGGTGTTGTGATGC	2096		
Db				
QY 1668	AATTAACATTTCAAAATAATTAAGATTTTTTTPAATAAATGCTTAAGAACGCTGATGTTG	1723		
Db				
QY 2097	AGCTAGTTTAAACAATAATCAACTTTTAACATATTAAATGGTAGTATCCGCTACTGTTG	2152		
Db				

RESULT 13
US-08-973-462-3
; Sequence 3, Application US/08973462B
; Patent No. 6191270

```

; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973.462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1891
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(1891)
US-08-973-462-3

```

Query Match	2.3%	Score 43.6	DB 4	Length 1891	
Best Local Similarity	43.9%	Pred. No. 0.058			
Matches 282	Conservative	0	Mismatches 354	Indels 6	Gaps 2
QY	75	AACGCGACGATATGCTAACCGGTAGTGGAGTCCCTGGTGCTCGACGAACCGTAACCTGC	134		
Db	505	AATCGTAGCTCCAACTGTTGAAGAAATTTAGCTCCCAAGTGTGTAGAAATGTGGCTCC	564		
QY	135	TGATGGTGCAGAGCTTGCAGCCGGAACAAATATAGGTCCCTGGAGCCGCTGCTTTGTATAGC	194		
Db	565	AAGTGTGTAAGAAAGTGTAGCTCCCAAGTGTGTAAGAAAGTGTAGCTGAAAATGTTGAAGA	624		
QY	195	GGGTCTTACTTTACAATATACCGGTGCTTTTACGGTTACTGATGCTGACGTAAGTGTTGC	254		
Db	625	AAGTGTAGCT---GAAATGTTGAAGAAATCGTAGCTCCCAAGTGTGTAAGAAAGTGTAGC	681		
QY	255	TGCATTAGATTTAAATAATTTGCAGCAGGTCTTTTTTTCAGTAACCTGGTGATATTTTCATT	314		
Db	682	TGAAATGTTTGAAGAAAGTGTAGCTGAAAATGTTTGAAGAAAGTGTAGCTGAAAATGTTGA	741		
QY	315	AGGTTCAAGTGTAGATACGGAGGAGCTATAAACTTGCAGTTAATATTGATGATGGTTT	374		
Db	742	AGAAAGTGTAGTGAAATGTTGAAG---AAAGTGTAGCTGAAAATGTTGAAGAAATCGT	798		
QY	375	AACCTTAACTTTAACAGGTACCGGTACTGCGAGCCTACGGTGCAAATCCTCGGTTGTTATT	434		
Db	799	AGCTCCAACTGTTGAAGAAAGTGTAGCTCCAACTGTTGAAGAAATTTAGCTCCAACCTGT	858		
QY	435	CCAAAGTGGACAGCTGCTCTTAATATACATATACCTGCTTTAGGTAAATACTCTAGG	494		
Db	859	TGAAGAAAGTGTAGCTCCAACTGTTGAAGAAATTTGATGTTCCAAAGTGTGGAAGAAAGTGT	918		
QY	495	TGAGGCGAATTCGCGGTTTGACTATTGCTTCAGATCCAGATGTATTAGGACCAATTAACGCT	554		
Db	919	AGCTCCAAGTGTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGT	978		
QY	555	TGCAGGAATATAGATGGAGGAGGTATAATACTGCACAATACAGATGCTGCCATTACGG	614		
Db	979	TGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGT	1038		
QY	615	AACAATAGGTAAATACTAATCCGCGACCTCAATTAAGCATTTGGAGCAAGTAGCGTTTCTCT	674		
Db	1039	AGCTGAAAATGTTGAAGAAATCGTAGCTCCCAAGTGTGGAAGAAATCGTAGCTCCAACCTGT	1098		
QY	675	TGAGGGGCGAGTTATTTAAAGCCACTACGACTAAGTTAAACAA	716		
Db	1099	TGAAGAAAGTGTGCTGAAACCGTTGCAACAAATTTATCAGA	1140		

RESULT 14
US-08-483-857-3

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 19:38:47 ; Search time 396 Seconds
(without alignments)
4870.603 Million cell updates/sec

Title: US-09-800-065-1
Perfect score: 1920
Sequence: 1 atggcgaatttctctaa.....cgttgagctaaacaatag 1920

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1920	100.0	1920	10	US-09-800-065-1
2	1920	100.0	5513	10	US-09-800-065-3
3	59.2	3.1	510	10	US-09-864-761-18737
4	57.6	3.0	3666	10	US-09-137-531-13
5	57.6	3.0	3666	10	US-09-137-531-14
6	57.6	3.0	4197	10	US-09-137-531-7
7	57.6	3.0	4197	10	US-09-137-531-8
8	50.4	2.6	2614	9	US-09-822-846-491
9	50.4	2.6	6604	10	US-09-880-107-1748
10	49.4	2.6	7104	10	US-09-815-242-4580
11	49.4	2.6	7107	10	US-09-815-242-8291
12	49.4	2.6	7434	10	US-09-815-242-4761
13	49.4	2.6	7437	10	US-09-815-242-8869
14	49.4	2.6	8155	7	US-08-781-986A-63
15	47.4	2.5	457	10	US-09-864-761-3110
16	46.8	2.4	5361	9	US-09-742-096-2
17	46.8	2.4	6152	9	US-09-742-096-1
18	46.6	2.4	172637	10	US-09-805-458A-3
19	45.6	2.4	1920	10	US-09-800-065-1

c 20	45.6	2.4	5513	10	US-09-800-065-3	Sequence 3, Appli
c 21	44	2.3	1635	10	US-09-864-761-20241	Sequence 20241, A
c 22	44	2.3	1973	10	US-09-864-761-3471	Sequence 3471, Ap
c 23	43.6	2.3	1891	9	US-09-742-096-4	Sequence 4, Appli
c 24	42.8	2.2	2694	10	US-09-815-242-8671	Sequence 8671, Ap
c 25	42.8	2.2	3343	10	US-09-815-242-4781	Sequence 4781, Ap
c 26	42.6	2.2	486	10	US-09-864-761-10113	Sequence 10113, A
c 27	42.4	2.2	512	10	US-09-864-761-25347	Sequence 25347, A
c 28	42.4	2.2	575	10	US-09-864-761-8628	Sequence 8628, Ap
c 29	41.8	2.2	1075	10	US-09-864-761-19241	Sequence 19241, A
c 30	41.8	2.2	1403	10	US-09-864-761-2513	Sequence 2513, Ap
c 31	41.6	2.2	3996	9	US-10-087-464-42	Sequence 42, Appli
c 32	41.2	2.1	439	10	US-09-864-761-20174	Sequence 20174, A
c 33	41.2	2.1	520	9	US-10-184-644-332	Sequence 332, App
c 34	41.2	2.1	520	9	US-10-184-634-332	Sequence 332, App
c 35	40.8	2.1	465	10	US-09-864-761-230	Sequence 230, App
c 36	40.8	2.1	31517	10	US-09-070-927A-180	Sequence 180, App
c 37	40.4	2.1	18692	9	US-10-092-154-1682	Sequence 1682, Ap
c 38	40.4	2.1	18692	10	US-09-764-847-1682	Sequence 1682, Ap
c 39	40.2	2.1	446	10	US-09-864-761-20699	Sequence 20699, A
c 40	40	2.1	2765	9	US-09-117-447-5	Sequence 5, Appli
c 41	40	2.1	4358	7	US-08-781-986A-454	Sequence 454, App
c 42	39.8	2.1	456	10	US-09-864-761-4249	Sequence 4249, Ap
c 43	39.8	2.1	593	9	US-09-783-252-20	Sequence 20, Appli
c 44	39.6	2.1	1141	9	US-10-222-162-49	Sequence 49, Appli
c 45	39.4	2.1	2050	9	US-10-086-510-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-800-065-1
; Sequence 1, Application US/098000065
; Patent No. US20020094552A1
; GENERAL INFORMATION:
; APPLICANT: Bouyer, Donald H
; APPLICANT: Croquet-Valdes, Patricia
; APPLICANT: Stenos, John
; APPLICANT: Walker, David H
; TITLE OF INVENTION: Rickettsia felis Outer Membrane Protein
; FILE REFERENCE: 026.00121
; CURRENT APPLICATION NUMBER: US/09/800,065
; CURRENT FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: US 60/187,323
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Rickettsia felis
US-09-800-065-1

Query Match	100.0%	Score 1920;	DB 10;	Length 1920;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1920;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGGCGAATATTCTCTAAATATTTCAAAAGCAATTCAAAAGGCTTTAAACTGCT	60	
Db	1	ATGGCGAATATTCTCTAAATATTTCAAAAGCAATTCAAAAGGCTTTAAACTGCT	60	
QY	61	TTATTACACACCTCAACCCAGCGATATGCTAACCGTAGTGGAGTCTCTGGTGTGCA	120	
Db	61	TTATTACACACCTCAACCCAGCGATATGCTAACCGTAGTGGAGTCTCTGGTGTGCA	120	
QY	121	AGAACCGTAACCTGCTGATGGTGCAGAGCTTCGAGCGGGAACAATATAGTCTCTGGAGCC	180	
Db	121	AGAACCGTAACCTGCTGATGGTGCAGAGCTTCGAGCGGGAACAATATAGTCTCTGGAGCC	180	
QY	181	GTGCTTTTGTAGCGGGTCTACTTTACAAATATACCGGTGCTTTTACGGTTACTGATGCT	240	
Db	181	GTGCTTTTGTAGCGGGTCTACTTTACAAATATACCGGTGCTTTTACGGTTACTGATGCT	240	

Applicant

```
QY 241 GACGTAAGTGTTCGTGCATTAAGATTTAAATAATTTTGACGAGGTCTTTTTTCAGTAAC 300
Db 241 GACGTAAGTGTTCGTGCATTAAGATTTAAATAATTTTGACGAGGTCTTTTTTCAGTAAC 300
QY 301 GGTGATATTTTCATTAAGTTCAGTGTAGATACGGAGGAGCTAATAAATCTGCAATTAAT 360
Db 301 GGTGATATTTTCATTAAGTTCAGTGTAGATACGGAGGAGCTAATAAATCTGCAATTAAT 360
QY 361 ATTGATGATGTTTAACTTTAACTTTAACTGAGTACCGGTACTGACGCTACGGTGCAAT 420
Db 361 ATTGATGATGTTTAACTTTAACTTTAACTGAGTACCGGTACTGACGCTACGGTGCAAT 420
QY 421 CCTCGGTTGTTTCCAAAGTGGACAGTGTCTGCTAATAATACATATACCTGCTTTAGGT 480
Db 421 CCTCGGTTGTTTCCAAAGTGGACAGTGTCTGCTAATAATACATATACCTGCTTTAGGT 480
QY 481 AATAAATCTCTAGTGTGAGGCAATGCCGTTTGAATTTGACTATTCCTTCAGATCCAGATGTA 540
Db 481 AATAAATCTCTAGTGTGAGGCAATGCCGTTTGAATTTGACTATTCCTTCAGATCCAGATGTA 540
QY 541 GGACCAATAACGCTTTCAGGAAATATAGATGGAGGAGTATATAAATCTGACAAATACAGAT 600
Db 541 GGACCAATAACGCTTTCAGGAAATATAGATGGAGGAGTATATAAATCTGACAAATACAGAT 600
QY 601 GCTGCCAATAACGCAATAGTAACTAATCCGGCAGCTCAATTAAGCAATACAGAT 660
Db 601 GCTGCCAATAACGCAATAGTAACTAATCCGGCAGCTCAATTAAGCAATACAGAT 660
QY 661 AGTACGCTTCTCTGTGAGGCGAGTATTAAAGCCACTACGACTAAGTTAAACAATGCA 720
Db 661 AGTACGCTTCTCTGTGAGGCGAGTATTAAAGCCACTACGACTAAGTTAAACAATGCA 720
QY 721 GCGCCGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 780
Db 721 GCGCCGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 780
QY 781 ACAGGCGGTGATGATGTAGTGTCTTAAATTTAAACGAGCGTGTAGTCAAGTAACGGA 840
Db 781 ACAGGCGGTGATGATGTAGTGTCTTAAATTTAAACGAGCGTGTAGTCAAGTAACGGA 840
QY 841 AATATAGTAAATACAAATTCATTAAGCACAATTAAGTGTAGGACGAGTACGGTACGTTA 900
Db 841 AATATAGTAAATACAAATTCATTAAGCACAATTAAGTGTAGGACGAGTACGGTACGTTA 900
QY 901 GGGGAGCGGTTTAAAGCCACTACGACTAAGTTAAACGAGCGTGTAGTCAAGTAACGGA 960
Db 901 GGGGAGCGGTTTAAAGCCACTACGACTAAGTTAAACGAGCGTGTAGTCAAGTAACGGA 960
QY 961 CTTACAATGCAGTATTAAACGAGTGTGTGATAACACACGAGCGGTGATAATGAGGT 1020
Db 961 CTTACAATGCAGTATTAAACGAGTGTGTGATAACACACGAGCGGTGATAATGAGGT 1020
QY 1021 GTCGTAATCTAAGTGGAGCATTGAGTCAAGTAAACGGGAATATAGTAAATACAAATTC 1080
Db 1021 GTCGTAATCTAAGTGGAGCATTGAGTCAAGTAAACGGGAATATAGTAAATACAAATTC 1080
QY 1081 TTAGCCACGATAAATATAGGACGAGTGTGGCTTACCTTGGATGAGCGGTATTAAAGCT 1140
Db 1081 TTAGCCACGATAAATATAGGACGAGTGTGGCTTACCTTGGATGAGCGGTATTAAAGCT 1140
QY 1141 ACTACGACTAAGTTAAACAGATGATGCTCAGTATTATTTTACGAATCCTGTAGTAGTA 1200
Db 1141 ACTACGACTAAGTTAAACAGATGATGCTCAGTATTATTTTACGAATCCTGTAGTAGTA 1200
QY 1201 ACCGGAGCAATAGATAATACCGGTAATGCGAATAAAGGTGTGTAATCTTTACCGGAGCA 1260
Db 1201 ACCGGAGCAATAGATAATACCGGTAATGCGAATAAAGGTGTGTAATCTTTACCGGAGCA 1260
QY 1261 AGTACGGTAACCGAATAATATAGGTAACAGCGCAGTATTAGCAGAGGTAAACGCTAGGAGCA 1320
Db 1261 AGTACGGTAACCGAATAATATAGGTAACAGCGCAGTATTAGCAGAGGTAAACGCTAGGAGCA 1320
```

```
QY 1321 GGTTCGTGCAATACAAAGCGGAGTAGTAAACGGAATGCAATAAACTTTAACGGATAAT 1380
Db 1321 GGTTCGTGCAATACAAAGCGGAGTAGTAAACGGAATGCAATAAACTTTAACGGATAAT 1380
QY 1381 GCGTCAGTAGTAACATTTACCGGTGATAGTACGGTAACAGGTAGTAGTGGTACAGAA 1440
Db 1381 GCGTCAGTAGTAACATTTACCGGTGATAGTACGGTAACAGGTAGTAGTGGTACAGAA 1440
QY 1441 CTCCTCCACACAGTGAATATAGGAGCAGGAATAACATTACGAGCCGAGGAAGCTAGCT 1500
Db 1441 CTCCTCCACACAGTGAATATAGGAGCAGGAATAACATTACGAGCCGAGGAAGCTAGCT 1500
QY 1501 GCGAATAATATAGATTTTGGAGCTGCGAGTAATTTAGAGTTTAAACGGTCTCTCCGCGTAAG 1560
Db 1501 GCGAATAATATAGATTTTGGAGCTGCGAGTAATTTAGAGTTTAAACGGTCTCTCCGCGTAAG 1560
QY 1561 AATTATTAACCTAATCGAATATAGCAACCGTAATTAATGCTACACCTAATATTAATGCT 1620
Db 1561 AATTATTAACCTAATCGAATATAGCAACCGTAATTAATGCTACACCTAATATTAATGCT 1620
QY 1621 GCTGGTACAGTATTGCAAAATGATTTAGTATAGGCACAGTTGACACAAATTAACATTCAA 1680
Db 1621 GCTGGTACAGTATTGCAAAATGATTTAGTATAGGCACAGTTGACACAAATTAACATTCAA 1680
QY 1681 AATAAATAAGATTTTGTAAATGCTTAAGACGCTGATGTTGATATATTAGAGCTCAG 1740
Db 1681 AATAAATAAGATTTTGTAAATGCTTAAGACGCTGATGTTGATATATTAGAGCTCAG 1740
QY 1741 GCGATTAGTTTAAAGGAGCAGCTTACGCTCTTTCTTAGCTAACGCTTAGCTACAGATG 1800
Db 1741 GCGATTAGTTTAAAGGAGCAGCTTACGCTCTTTCTTAGCTAACGCTTAGCTACAGATG 1800
QY 1801 ATAGAGTTATCACTCTTAAAAATCATTTACCGGCTCTTGTCTAACGGTGGTGGAGTTAA 1860
Db 1801 ATAGAGTTATCACTCTTAAAAATCATTTACCGGCTCTTGTCTAACGGTGGTGGAGTTAA 1860
QY 1861 TTCTTTTGGAGTCCAAACAACTTTATGACCGTACAAAGGTGACGTTGGAGCTTAAACATAG 1920
Db 1861 TTCTTTTGGAGTCCAAACAACTTTATGACCGTACAAAGGTGACGTTGGAGCTTAAACATAG 1920
```

RESULT 2

US-09-800-065-3

; Sequence 3, Application US/09800065

; Patent No. US20020094552A1

; GENERAL INFORMATION:

; APPLICANT: Bouyer, Donald H

; APPLICANT: Crocquet-Valdes, Patricia

; APPLICANT: Stenos, John

; APPLICANT: Walker, David H

; TITLE OF INVENTION: Rickettsia felis Outer Membrane Protein

; FILE REFERENCE: 026.00121

; CURRENT APPLICATION NUMBER: US/09/800,065

; CURRENT FILING DATE: 2001-03-06

; PRIOR APPLICATION NUMBER: US 60/187,323

; PRIOR FILING DATE: 2000-03-06

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 5513

; TYPE: DNA

; ORGANISM: Rickettsia felis

US-09-800-065-3

Query Match 100.0%; Score 1920; DB 10; Length 5513;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGAATATTTCTCTAAATTTATTTCAAAAAGCAATTTCAAAAAGGTCTTTAAACTGCT 60

Db 478 ATGGCGAATATTTCTCTAAATTTATTTCAAAAAGCAATTTCAAAAAGGTCTTTAAACTGCT 537

QY 61 TTATTCACCACCTCAACCCGAGGATAATGCTAACCGGTAGTGGAGTCTCTTGGTGTGCA 120

Db 538 TTAATTCACACCTCAACCGCAGCGAATAATGCTAAACCGGTAGTGGAGTCTTGGTGCTGCA 597
 Qy 121 AGAACCGTAACCTGCTGATGCTGAGAGCTTGCAGCGCGAACAATAATATAGTCTCGGAGCC 180
 Db 598 AGAACCGTAACCTGCTGATGCTGAGAGCTTGCAGCGCGAACAATAATATAGTCTCGGAGCC 657
 Qy 181 GGTGCTTTTGTAGCGGGTCTACTTTACAATAATACCGGTGCTTTTACGGTTACTGATGCT 240
 Db 658 GGTGCTTTTGTAGCGGGTCTACTTTACAATAATACCGGTGCTTTTACGGTTACTGATGCT 717
 Qy 241 GACGTAAGTGTTCGCGCATTAATTAATAATTTTGCAGCAGGCTTTTTCAGTAAC 300
 Db 718 GACGTAAGTGTTCGCGCATTAATTAATAATTTTGCAGCAGGCTTTTTCAGTAAC 777
 Qy 301 GGTGATATTCATTAAGTTCAGTGTAGATACGGGAGGAGCTAATAAATCTGCGATTAAT 360
 Db 778 GGTGATATTCATTAAGTTCAGTGTAGATACGGGAGGAGCTAATAAATCTGCGATTAAT 837
 Qy 361 ATTGATGATGGTTTAACCTTAACCTTAACAGGTACCGGTACTGCGAGCCTACGGTCAAA 420
 Db 838 ATTGATGATGGTTTAACCTTAACCTTAACAGGTACCGGTACTGCGAGCCTACGGTCAAA 897
 Qy 421 CTTGCGTGTATTCGAAGTGGACAAGCTGCTGCTAATAATATACATATACTGCTTTAGGT 480
 Db 898 CTTGCGTGTATTCGAAGTGGACAAGCTGCTGCTAATAATATACATATACTGCTTTAGGT 957
 Qy 481 AATATACTCTAGTGGAGCGAATCGCGTTTGACTATTTGCTTCAGTCCAGATGATTA 540
 Db 958 AATATACTCTAGTGGAGCGAATCGCGTTTGACTATTTGCTTCAGTCCAGATGATTA 1017
 Qy 541 GCACCAATAACCTTGCAGGAATAATAGATGAGGAGGTATAAATACGCACAAATACAGAT 600
 Db 1018 GGACCAATAACCTTGCAGGAATAATAGATGAGGAGGTATAAATACGCACAAATACAGAT 1077
 Qy 601 GGTGCATTAAACGGGAATAATAGTAACTAATCCGCGAGCTCAATAAGCAATGCA 720
 Db 1078 GGTGCATTAAACGGGAATAATAGTAACTAATCCGCGAGCTCAATAAGCAATGCA 1137
 Qy 661 AGTACGCTTCTCTGGAGGGGAGTATTAAGCCACTACGACTAAGTTAACAATGCA 720
 Db 1138 AGTACGCTTCTCTGGAGGGGAGTATTAAGCCACTACGACTAAGTTAACAATGCA 1197
 Qy 721 GCGCCGCTATTAACACTTACAATGCAATGCAATTAACAGGCTGCTGCTGATACACC 780
 Db 1198 GCGCCGCTATTAACACTTACAATGCAATGCAATTAACAGGCTGCTGCTGATACACC 1257
 Qy 781 ACAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 Db 1258 ACAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1317
 Qy 841 AATATAGGTAATACAAATTCATAGCGACAATAAGTGTAGGAGCAGGTACGCTAGCTTA 900
 Db 1318 AATATAGGTAATACAAATTCATAGCGACAATAAGTGTAGGAGCAGGTACGCTAGCTTA 1377
 Qy 901 GGGGAGCGGTATTAAGCCACTACGACTAAGTTAACGAATACGCGCTGCTATTAACA 960
 Db 1378 GGGGAGCGGTATTAAGCCACTACGACTAAGTTAACGAATACGCGCTGCTATTAACA 1437
 Qy 961 CTTACAAATGCAATTAACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 Db 1438 CTTACAAATGCAATTAACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1497
 Qy 1021 GTCGTAATCTAGTGGAGCATTAAGTCAAGTAACCGGAATATAGTAAATCAATTC 1080
 Db 1498 GTCGTAATCTAGTGGAGCATTAAGTCAAGTAACCGGAATATAGTAAATCAATTC 1557
 Qy 1081 TTAGCCACGATAAATATAGGAGCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
 Db 1558 TTAGCCACGATAAATATAGGAGCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1617
 Qy 1141 ACTACGACTAAGTTAACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200

Db 1618 ACTAGGCTAAGTTAAACAGATGATCGGTCAGTATTAAATATTTACGAATCCTGTAGTGA 1677
 Qy 1201 ACCGAGCAATAGATAATACCGGTAAATGCCAATAAAGGTGGTAATCTTTTACCGGAGCA 1260
 Db 1678 ACCGAGCAATAGATAATACCGGTAAATGCCAATAAAGGTGGTAATCTTTTACCGGAGCA 1737
 Qy 1261 AGTAGGTAACCGAATAATATAGTAAACACGCGAGTATTAGCAGAGGTAAAGCGTAGGAGCA 1320
 Db 1738 AGTAGGTAACCGAATAATATAGTAAACACGCGAGTATTAGCAGAGGTAAAGCGTAGGAGCA 1797
 Qy 1321 GGTTCGTCGCAATACAAAGCGGAGTAGTAAAGCGAATGCAATAAATTAACGATTAAT 1380
 Db 1798 GGTTCGTCGCAATACAAAGCGGAGTAGTAAAGCGAATGCAATAAATTAACGATTAAT 1857
 Qy 1381 GCGTCAGTAGTAACATTTACCGGTGATAGTAGTAACAGGTAGTATAGTGGTACAGAA 1440
 Db 1858 GCGTCAGTAGTAACATTTACCGGTGATAGTAGTAACAGGTAGTATAGTGGTACAGAA 1917
 Qy 1441 CTCTTCGCAACAGTGAATATAGGAGCAGGAATAACATTACGAGCGGAGGAAGCCTAGCT 1500
 Db 1918 CTCTTCGCAACAGTGAATATAGGAGCAGGAATAACATTACGAGCGGAGGAAGCCTAGCT 1977
 Qy 1501 GCGAATAATATAGATTTTGGAGCTCGGAGTAATTTAGAGTTTAAAGGCTTCCCGGTAA 1560
 Db 1978 GCGAATAATATAGATTTTGGAGCTCGGAGTAATTTAGAGTTTAAAGGCTTCCCGGTAA 2037
 Qy 1561 AATTAACTTAATCGGAATATAGCAACGCTAATAATGCTACACCTTAATATTAATGCT 1620
 Db 2038 AATTAACTTAATCGGAATATAGCAACGCTAATAATGCTACACCTTAATATTAATGCT 2097
 Qy 1621 GCTGTACAGTGAATGCAAAATGATTTAGTAGTAGGACAGCTTGCACAAATTAACATTC 1680
 Db 2098 GCTGTACAGTGAATGCAAAATGATTTAGTAGTAGGACAGCTTGCACAAATTAACATTC 2157
 Qy 1681 AATAATAGATTTTGTATAAATGCTAAGACGCTGATGTTGATATATTAACGCTCAG 1740
 Db 2158 AATAATAGATTTTGTATAAATGCTAAGACGCTGATGTTGATATATTAACGCTCAG 2217
 Qy 1741 GCGATTAGTTTAAAGGAGCAGCTTACAGCTCTTTTCTTACGTAACGCTTACGATG 1800
 Db 2218 GCGATTAGTTTAAAGGAGCAGCTTACAGCTCTTTTCTTACGTAACGCTTACGATG 2277
 Qy 1801 ATAGAGTTATCACTCTTAAAAATCAATTTACCGGCTTCTGCTACGCTGCTGCTGCTGCTGCTGCT 1860
 Db 2278 ATAGAGTTATCACTCTTAAAAATCAATTTACCGGCTTCTGCTACGCTGCTGCTGCTGCTGCT 2337
 Qy 1861 TTCTTTTGGAGTCCACAAACCTTATGACCGTCAAGGTCAGGCTGAGCTAAGCAATAG 1920
 Db 2338 TTCTTTTGGAGTCCACAAACCTTATGACCGTCAAGGTCAGGCTGAGCTAAGCAATAG 2397

RESULT 3

US-09-864-761-18737/c
 ; Sequence 18737, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aemica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04

```

; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18737
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035457.11
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.85
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
; OTHER INFORMATION: NT HIT: U01287.1, EVALUATE 1.30e+00
US-09-864-761-18737

Query Match      3.1%; Score 59.2; DB 10; Length 510;
Best Local Similarity 46.1%; Pred. No. 0.0007;
Matches 199; Conservative 0; Mismatches 233; Indels 0; Gaps 0;

QY 104 GAGTCCTTGGTGCAGACCGCTAACTGCTGATGCTGACAGCTTGCAGCCGGAACAA 163
DB 510 GTGTTGATGCTGATGCTGAGGACCGCTGCTGAAGCAGGTGTTGATGCTGCTGAAAAG 451

QY 164 ATATAGTCTTGGACCGCGTCTTTTACGGGTTCTACTTTACAATATACCGTCTT 223
DB 450 GTGTTGCTGAGGTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 391

QY 224 TTACGGTTACTGATGCTGACGTAAGTCTTCGTCATTAGATTTAAATTTTGCAGCAG 283
DB 390 ATGCAGGTATTTGTTGTTGCTGAGATGTTGATGCTGATGTTGAAGCAGGTGCTGAAAAG 331

QY 284 GTCTTTTTCAGTAACCTGGGATATTCATTAGTTTCAGTGGTAGATACGGGAGGAGCTA 343
DB 330 GTGTTGATGCTGAGGTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 271

QY 344 ATAACTTGCAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 403
DB 270 CAGATGTTGATGCTGAGATGTTGAAGCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 211

QY 404 CAGCCTACGCTGCAATCTCGCTTGTATTCCAAAGTGGACAGCTGCTGCTGCTGCTGCT 463
DB 210 CAAGTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 151

```

```

QY 464 CATATACCTGCTTTAGTAAATATACTCTAGGTGGAGCAATGCCGGTTTGACTATTGCTT 523
DB 150 TTGATGCAGATGTTGAAGCGGGTGTGTGATGCTGATGCAGGTGCTGAAGCAGGTATTGCTG 91
QY 524 CAGATCCAGATG 535
DB 90 CTGATACAGATG 79

RESULT 4
US-09-137-531-13
; Sequence 13, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EP0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/692,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-137-531-13

Query Match      3.0%; Score 57.6; DB 10; Length 3666;
Best Local Similarity 47.0%; Pred. No. 0.0042;
Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;

QY 116 CTGCAAGAACCGTAACCTGCTGATGGTGCAGAGCTTGCAGCCGGGAACAAATATAGTCTG 175
DB 2735 CTTTAGGAATTTTCATTAGCTGATGCAGATCTTAATGTAAGTGCACAACTGTTGATACTG 2794

QY 176 GAGCCGGTGTCTTTTGTAGCCGGTCTACTTTACAATATACCGGTGCTTTTACGGTTACTG 235
DB 2795 CAACTGTTTCATTAAAG-----ATAGTCAAAATTAATTCATTATCTCTTACATTAGTTG 2848

QY 236 ATGCTGACGTAAGTGTTCGTGCATTAGATTTAATAATTTTGCAGCAGGTCTTTTTCAG 295
DB 2849 AACTGCTGTAATACAGGTTGATTTGCTACAACCTGTTCAAGCTGGTACATTATCTTCTT 2908

QY 296 TAACCTGGTATATTTTCATTAGTTCAGTGTAGATACGGGAGGAGCTAATAAATCTTCAG 355
DB 2909 TAACCTGCTGATACATAACAGTTACTTATGCAGATGCTAAATAATGCTGCAGGTGTTGCTG 2968

QY 356 TTAATATTGATGATGTTTAACTTAACTTTACAGGTACCGGTACTGACGCTACCGGTG 415
DB 2969 AAAATATTACTGCTAGCTAAACATTAAAGAAAAC---TACTGGAGCAATTTACTTCTGATA 3025

QY 416 CAAATCCTGCTGTTTATTCCAAAGTGGACCAAGCTGCTGCTAATAATACATATACCTGCTT 475
DB 3026 CATTACACAAGGTGATTACCATCAGCAGCTACACAGCTGAATATATCTTCTTAATCAA 3085

QY 476 TAGGTAATACTAGTGGAGCGAATGCCGGTTTGACTATTGCTTCAGATCCAGATG 535
DB 3086 TTGCTCAGATTATACATTTGCAACAGGTGAAGTTCACCTTTTAAATATTGATAATGCTG 3145

QY 536 TATTAGGACCAATAACCGTTGCAGGAAATATAGATGAGGAGGTATAATAACTGACAATA 595

```

Query Match	3.0%;	Score	57.6;	DB	10;	Length	3666;
Best Local Similarity	47.0%;	Pred. No.	0.0042;				
Matches	285;	Conservative	0;	Mismatches	309;	Indels	12;
Gaps							

QY	116	CTGCAAGAACGGTAACGTGTCAGTGGTGCAGAGCTTGCAGCGCGGAACAAATATAGTCTCTG	175
DB	2735	CTTTAGGAATTTCAATTAGCTGATGCAGATCTTAATGTAAGTGCACCAACTGTTGATACTG	2794
QY	176	GAGCCGCTGCTTTGTGTAGCGGTTCTACTTTACAATATACCGGGTCTTTTACGGTTACTG	235
DB	2795	CAACTGTTTTCATTAANAAG-----ATAGTGCAAATAATTCATTCTCTTACATTAGTGT	2848
QY	236	ATGCTGACGTAAGTGTCGTGCATTAGATTTAAATAATTTTGCAGCAGGTCCTTTTTTCAG	295
DB	2849	AACTGGTGCTAATACAGGTGTAATTTGCTACAACATGTTCAAGCTGGTACATTATCTCTT	2908
QY	296	TAACCTGGTGATATTTCATTAGTTTCAGTGGTAGATACGGGAGGACTAATAAACTTTGCAG	355
DB	2909	TAACCTGGTGTCATTAACAGTTACTTATGCAGATGCTAAAANTGCTCCAGGTGTGCTG	2968
QY	356	TTAATATTGATGATGGTTTAACTTTAAACAGGTACCGGTACTGCAGCCCTACGGTG	415
DB	2969	AAAATATTACTGCTAGCGTACATTAAGAANAAC---TACTTGGACCAATTACTTCTGATA	3025
QY	416	CAATCTCGCTTGTATTCCAAGGTGGACAGCTGCTGCTTAATAATAATATATACTGCTT	475

Db	3026	CATTTCACAAGAGTGTAATTACCATCAGCAGCTACAGCAGCTGAATATATCTTCAAAATCAA	3088
Qy	476	TAGTGAATATAACTCTAGTGTGGAGGCAATGCCGGTTTGACTATTGCTTCAGATCCAGATG	535
Db	3086	TTGCTGCAGATTATACATTTGCAACAGGTGAAGGATTCACCTTTAAATATTGATAATGCTG	3145
Qy	536	TATTAGACCAATAAAGCTTTGACGAGGAATATAGATGGAGGAGGTATTAATAACTGCACAATA	595
Db	3146	GTGCTCAAGTAATAACTTATGACAGGTAAAGAGGTGCAACAAGGT---GTAGCTGATGCTA	3202
Qy	596	CAGATGCTGCATTAAACGGAACAATAAGGTAAATACCTAATCCGGCAGCTCAAAATAGCATTG	655
Db	3203	TCAATGCTACATTTGCGAGGTACTGCACTGTTCTCGGAGACAAAGTAGTTATTAAATCAG	3262
Qy	656	GAGCAAGTACGCTTCTCTTTGGAGGGCAGTTATTAAAGCCACCTACGACTAAGTTAAACAA	715
Db	3263	CTACACAGGTGTTGCTTCTGAAGTTGAAGTTACATCTCTCTGTTAAATCAAGTATTAA	3322
Qy	716	ATGCAG 721	
Db	3323	ATGCAG 3328	
RESULT 6			
US-09-137-531-7			
: Sequence 7, Application US/09137531			
: Patent No. US20020048816A1			
: GENERAL INFORMATION:			
: APPLICANT:			
: TITLE OF INVENTION: Expression of surface layer proteins			
: NUMBER OF SEQUENCES: 25			
: COMPUTER READABLE FORM:			
: MEDIUM TYPE: Floppy disk			
: COMPUTER: IBM PC compatible			
: OPERATING SYSTEM: PC-DOS/MS-DOS			
: SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)			
: CURRENT APPLICATION DATA:			
: APPLICATION NUMBER: US/09/137,531			
: FILING DATE:			
: CLASSIFICATION:			
: PRIOR APPLICATION DATA:			
: APPLICATION NUMBER: 08/682,517			
: FILING DATE:			
: INFORMATION FOR SEQ ID NO: 7:			
: SEQUENCE CHARACTERISTICS:			
: LENGTH: 4197 base pairs			
: TYPE: nucleic acid			
: STRANDEDNESS: single			
: TOPOLOGY: linear			
: MOLECULE TYPE: DNA (genomic)			
US-09-137-531-7			
Query Match 3.0%; Score 57.6; DB 10; Length 4197;			
Best Local Similarity 47.0%; Pred. No. 0.0044;			
Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps			
Qy	116	CTGCAAGAACCGTAACCTGCTGTGGTGACAGAGCTTCGACCGGAAACAATATAGTCCCTG	175
Db	2919	CTTTAGANAATTCATTAGCTGATGTCAGATCTTAACTGTAAGTCAACAACACTGTTGATAGT	2978
Qy	176	GAGCCGGTGCTTTTGTAGCCGGGTTCTACTTTACAATATACCGGGTGCTTTTACGGTTACTG	235
Db	2979	CAACTGTTTCATTAAAAG-----ATAGTGAACAATAATTCATTATCTCTTACATTAGTGTG	3032
--Qy	236	ATGCTGACGTAAGTGTTCGTGCATTAGATTAAATAATTTTTCAGCAGGCTCTTTTTCAG	295
Db	3033	AAACTGGTGCTTAATACAGGTGATTTTGCTACAACTGTTCAAGCTGCTACATTATCTTCTT	3092
Qy	296	TAACTGGTGATTTTTCATTAGTTTCACTGGTGTAGATACGGGAGGAGCTAATAAATCTGCAG	355
Db	3093	TAACTGGTGCTGATTAATTAACAGTTACTTATGCAATGCTAAAATGCTGACGGTGTGCTG	3155
Qy	356	TTAATATTGATGATGGTGTTTAACTTTAACTTTAAACAGGTACCCGGTACTGCCAAGCTACGGTG	415

Db 3153 AAAATATTACTGCTAGCGTAACATTAAAGAAAC---TACTGGAGCAATTAATCTCTGATA 3209
Qy 416 CAAATCTCGTGTGTTATTCCAAAGTGGACAGCTGCTGCTAAATATACATATATCTGCT 475
Db 3210 CATTTACACAGGTGTATTACCATCAGCAGCTACAGAGCTGAATATATCTTCTAAATCAA 3269
Qy 476 TAGGTAATATACTCTAGGTGGAGGGAATGCGGTTTGTGACTATTGCTTCAGATCCAGATG 535
Db 3270 TTGCTGCAGATTATACATTTGCAACAGGTGAAGGATTCACITTAATATTGTAATGCTG 3329
Qy 536 TATTAGGACCATACGCTTCGACGAAATATAGATGGAGGAGGTATATACTGACAAATA 595
Db 3330 GTGCTCAAGTAATTAACCTTAGCAGGTAAAGGTCACAAGGT---GTAGCTGATGCTA 3386
Qy 596 CAGATGCTCCATTAAACGGAACAATAGGTAACTACTAATCCCGCAGCTCAAAATACGATTG 655
Db 3387 TCAATGCTACATTTGCCAGTACTGCACTGTTCTGGAGACAAGTAGTATTAAATCAG 3446
Qy 656 GAGCAAGTACGCTTCTCTGGAGGGCAGTTATTAAAGCCACTACGACTAAGTTAAACAA 715
Db 3447 CTACACAGAGGTGTTGGTTCTGGAAGTTGAAGTTACATTTCTTCTGTTAATCAAGTATTAA 3506
Qy 716 ATGCAG 721
Db 3507 ATGCAG 3512

RESULT 7

US-09-137-531-8
; Sequence 8, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus sphaericus
; INDIVIDUAL ISOLATE: P-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..3850
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 185..3850
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95..184

US-09-137-531-8

Query Match 3.0%; Score 57.6; DB 10; Length 4197;
Best Local Similarity 47.0%; Pred. No. 0.0044;
Matches 285; Conservative 0; Mismatches 309; Indels 12; Gaps 3;

Qy 116 CTGAAGAACCCTAACTGCTGATGTCAGAGCTTGCAGCGGACCAACAATATAGTCTCTG 175
Db 2919 CTTTAGGAATTTCAATGCTGATGCAGATCTTAATGTAAAGTGCACAACTCTTGTATAGTG 2978
Qy 176 GAGCCGCTGCTTTTGTAGCGGTTCTACTTTTCAATATACCGGTGCTTTTACGGTTACTG 235
Db 2979 CAACCTGTTTCAATTAAG-ATAGTGCAAATTAATTCATTATCTTACATTAAGTTG 3032
Qy 236 ATGCTGACGTAAGTGTGCTGCAATAGATTAATAATAATTTTGCAGCAGGTCTTTTTCAG 295
Db 3033 AAATCTGCTGCTAATACAGGTGTTTGTACAACCTGTTCAAGCTGCTACATTAATCTTCT 3092
Qy 296 TAACTGCTGATATTTCAATAGTTTCAGTGTAGATAGCGGAGGAGCTAAATAAATTCAG 355
Db 3093 TAACTGCTGCTAATACAGGTGTTTATGCGATGCTTAAATAATGCTGAGGTGTTGCTG 3152
Qy 356 TTAATATGATGATGTTTAACTTAACTTTTACAGAGTACCGGTACTGCGCCTACGGTG 415
Db 3153 AAAATATTACTGCTAGCGTAACATTAAAGAAAC---TACTGGAGCAATTAATCTTCTGATA 3209
Qy 416 CAAATCTCGCTGTTGTTTCAAGGTGGACAGCTGCTGCTAATAATACATATACATGCTT 475
Db 3210 CATTTACACAGGTGTTATACCATCAGCAGCTACAGCAGCTGAATATACATTAATCAA 3269
Qy 476 TAGTAATATAACTCTAGTGGAGCGAATCCCGTTTTCAGCTATTGCTTCAGATCCAGATG 535
Db 3270 TTGCTGCAGATTATACATTTTGCACAGGTGAAGGATTCACITTAATATTGATAATGCTG 3329
Qy 536 TATTAGGACCATACCGCTTGCAGGAATATAGATGGAGGAGGTATTAATACATGACAATA 595
Db 3330 GTGCTCAAGTAATTAACCTTAGCAGGTAAAGGTCACAAGGT---GTAGCTGATGCTA 3386
Qy 596 CAGATGCTCCATTAAACGGAACAATAGGTAACTACTAATCCCGCAGCTCAAAATACGATTG 655
Db 3387 TCAATGCTACATTTGCAGGTACTGCACTGTTCTGGAGACAAGTAGTATTAAATCAG 3446
Qy 656 GAGCAAGTACGCTTCTCTGGAGGGCAGTTATTAAAGCCACTACGACTAAGTTAAACAA 715
Db 3447 CTACACAGAGGTGTTGGTTCTGGAAGTTGAAGTTACATTTCTTCTGTTAATCAAGTATTAA 3506
Qy 716 ATGCAG 721
Db 3507 ATGCAG 3512

RESULT 8

US-09-822-846-491/c
; Sequence 491, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6400
; CURRENT APPLICATION NUMBER: US/09/822,846
; CURRENT FILING DATE: 2001-03-29

```

; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 491
; LENGTH: 2614
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-846-491

Query Match      2.6%; Score 50.4; DB 9; Length 2614;
Best Local Similarity 47.7%; Pred. No. 0.15;
Matches 147; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 125 CGTAACCTGCTGATGCTGCAGCTTCGACCGGAGCAAAATAGTCTCGAGCCGGT 184
Db 2295 CTGTTGCTGCTGCTGGGTCTGTAAGCCCTGGCGCTGGAACTGGGCTGGAGTGGG 2236

QY 185 CTTTGTAGCGGGTCTACTTTTACAATATACCGGTGCTTTTACGGTTACTGATGCTGACG 244
Db 2235 CTGGGGTTGGGAGGAGCGGCTGTGCTGCTGTTGCTGCTGGTGTGTTGCTG 2176

QY 245 TAAAGTGTGCTGATAGATTAATAAATTTTCAGCAGGCTCTTTTTCAGTAACCTGGTG 304
Db 2175 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2116

QY 305 ATATTTCATTAGGTCAGTGTAGATACGGGAGGAGCTAAATAAACTTGCAGTTAATATG 364
Db 2115 CTGCGCAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2056

QY 365 ATGATGGTTTAACTTTTAAACAGGTACCGGTACTGCAGCCTACGGTGCAAAATCTTG 424
Db 2055 CTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1996

QY 425 CATTGTTA 432
Db 1995 CTCAGGTA 1988

RESULT 9
US-09-880-107-1748/c
; Sequence 1748, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1748
; LENGTH: 6604
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D83783
US-09-880-107-1748

Query Match      2.6%; Score 50.4; DB 10; Length 6604;
Best Local Similarity 47.7%; Pred. No. 0.23;
Matches 147; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 125 CGTAACCTGCTGATGCTGCAGAGCTTCGACCGGAGCAAAATAGTCTCTGGAGCCGGTG 184
Db 6291 CTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6232

; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 491
; LENGTH: 2614
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-846-491

Query Match      2.6%; Score 50.4; DB 9; Length 2614;
Best Local Similarity 47.7%; Pred. No. 0.15;
Matches 147; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 125 CGTAACCTGCTGATGCTGCAGCTTCGACCGGAGCAAAATAGTCTCGAGCCGGT 184
Db 2295 CTGTTGCTGCTGCTGGGTCTGTAAGCCCTGGCGCTGGAACTGGGCTGGAGTGGG 2236

QY 185 CTTTGTAGCGGGTCTACTTTTACAATATACCGGTGCTTTTACGGTTACTGATGCTGACG 244
Db 2235 CTGGGGTTGGGAGGAGCGGCTGTGCTGCTGTTGCTGCTGGTGTGTTGCTG 2176

QY 245 TAAAGTGTGCTGATAGATTAATAAATTTTCAGCAGGCTCTTTTTCAGTAACCTGGTG 304
Db 2175 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2116

QY 305 ATATTTCATTAGGTCAGTGTAGATACGGGAGGAGCTAAATAAACTTGCAGTTAATATG 364
Db 2115 CTGCGCAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2056

QY 365 ATGATGGTTTAACTTTTAAACAGGTACCGGTACTGCAGCCTACGGTGCAAAATCTTG 424
Db 2055 CTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1996

QY 425 CATTGTTA 432
Db 1995 CTCAGGTA 1988

RESULT 10
US-09-815-242-4580
; Sequence 4580, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4580
; LENGTH: 7104
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4580

Query Match      2.6%; Score 49.4; DB 10; Length 7104;
Best Local Similarity 44.2%; Pred. No. 0.4;
Matches 249; Conservative 0; Mismatches 311; Indels 3; Gaps 1;

QY 725 CGGTATTAACACTTACAAATGCAATGCAAGTATTAACAGGTGCTGTGATAACACCACAG 784
Db 3809 CTGTGCAACTCAAAATCAAGCAATGATAATACAACACTGGTGTCTACAACTGAAGAAAAA 3868

QY 785 GCGGTGATGATGATGCTGCTTAAATTTAAACGAGCGGTGTGAGTCAAGTAACTGGAATA 844
Db 3869 ATGCAGCAAAAGATTTTAGTTTTAAAGCTAAAGAAAAACGATATCAAGATATCTTAAATG 3928

```


RESULT 13

US-09-815-242-8869

; Sequence 8869, Application US/09815242

; Patent No. US20070061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

Db 4700 TTAGCAGCATATGAAGAAGT 4722

RESULT 14

US-08-781-986A-63
; Sequence 63, Application US/08781986A
; Publication No. US20030054436A1

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 5255

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/781,986A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Benson, Bob

; REGISTRATION NUMBER: 30,446

; REFERENCE/DOCKET NUMBER: PB248PP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 63:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 8155 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-08-781-986A-63

Query Match 2.6%; Score 49.4; DB 7; Length 8155;

Best Local Similarity 44.2%; Pred. No. 0.43;

Matches 249; Conservative 0; Mismatches 311; Indels 3; Gaps 1;

QY 725 CGGTATTAAACACTTACAATGCAATGCAATTAACAGGTGCTGTTGATTAACACAGT 784

Db 6152 CTGTGACAACTCAAAATCAAGCAATTTGATTAACAACTGGTGTACAACTGAAGAGAAA 6211

QY 785 GCGGTGATGATGAGGTGCTTAAATTTAAACGAGCGTTGAGTCAAGTAACTGGAATA 844

Db 6212 ATGCAGCAAGATTGATTTAAAGCTTAAGAAAAGCGTATCAAGATATCTTAAGT 6271

QY 845 TAGTAATACAAATTCATTAGCGCAATAAGTGTAGGAGCAGGTACGCGTACGTTAGGG 904

Db 6272 CACAACAACATGATGTTACGCAAAATTAAGATCAACAGTGTGCTGATTTCAAGTA 6331

QY 905 GACGGTTTAAAGCCACTACGACTAAGTTAAGCAATG---CAGCGTCGGTATTACAC 961

Db 6332 TTACTGCAGATACAACAATTAAGATTTGCGAAAGATGATTAAGCAACAAAAGCAACG 6391

QY 962 TTCAAAATGCGATTTAAACAGGTGCTGTTGATTAACACACGAGCGGTGATAGTGTG 1021

Db 6392 AAAAAAGCGCTTATTTGCAAACTCCAGTGGGACTCTGAAGAAAAAGCAACAA 6451

QY 1022 TCGTAATCTAAGTGGAGCATTTAGTCAAGTAACCGGGAATATAGGTAATACAAATTCAT 1081

Db 6452 ATCAACAAGTAGCGCAATTAACACAAGGTATCAAAATATTGAAAAATGCACAGTCAA 6511

QY 1082 TAGCCAGCATAAATATAGGAGCAGGTGTGGCTACTCTGGATGGAGCGGTTATTAAAGCTA 1141
Db 6512 TCGATGATGTAACACACTGCAAAAGATAATGCAATTCAGCAATTCAGCAATTCAGCAT 6571
QY 1142 CTACGACTAAGTTAACAGATGATCGGCAGTATTAAATTTACGAATCCTCTAGTAGTAA 1201
Db 6572 CAACAGATGTTAAACGAATGCAAGAGCGGAATTCCTAACTGAAATGCAAAATAAAAATAA 6631
QY 1202 CCGGAGCAATAGATAATACCGGTAAATGCCAATAAAGGTGTGGTAACTCTTTACCGGAGCAA 1261
Db 6632 CTGAATACITTAATAATAGAGACTACTAATGAAGAAAAAGTAAAGTATTTGGACCAG 6691
QY 1262 GTACGGTAACGCAATAATATAGGT 1284
Db 6692 TTAGAGCAGCATATGAAGAAGT 6714

RESULT 15

US-09-864-761-3110/c

; Sequence 3110, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864, 761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 3110

; LENGTH: 457

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:
; OTHER INFORMATION: MAP TO AL035457.11
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.85
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
US-09-864-761-3110

Query Match 2.5%; Score 47.4; DB 10; Length 457;
Best Local Similarity 52.2%; Pred. No. 0.3;
Matches 105; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 103 GGAGTCCTGGTGGTCCGAAGACCGTAAGTGGTGCAGAGCTTGCAGCCGGAACA 162
DB 103 GGAGTCCTGGTGGTCCGAAGACCGTAAGTGGTGCAGAGCTTGCAGCCGGAACA 162
QY 201 GGTGTTGATGCTGATGCAGGTACCGATGCTGAAGCAGGTGTTGATGCAGGTGCTGAAAA 142
DB 201 GGTGTTGATGCTGATGCAGGTACCGATGCTGAAGCAGGTGTTGATGCAGGTGCTGAAAA 142
QY 163 AATATAGTCTCGAGCGCGTCTTTTGTAGCGGTTCTTACTTTACAATATATACCGTGCT 222
DB 163 AATATAGTCTCGAGCGCGTCTTTTGTAGCGGTTCTTACTTTACAATATATACCGTGCT 222
QY 141 GGTGTTGGTGCAGTGCAGATGCTGATGCTGCTAGTAAATGCTGATGCAGCGCTGATGCT 82
DB 141 GGTGTTGGTGCAGTGCAGATGCTGATGCTGCTAGTAAATGCTGATGCAGCGCTGATGCT 82
QY 223 TTTACGGTTACTGATGCACGTAACTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 282
DB 223 TTTACGGTTACTGATGCACGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 282
QY 81 GATCAGGTATTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 22
DB 81 GATCAGGTATTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 22
QY 283 GGTCTTTTTCAGTAACTGGT 303
DB 283 GGTCTTTTTCAGTAACTGGT 303
QY 21 GGTGTTGATGCAGTGCAGAT 1
DB 21 GGTGTTGATGCAGTGCAGAT 1

Search completed: April 17, 2003, 19:48:13
Job time : 558 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 13:33:35 ; Search time 8577 Seconds
(without alignments)
5628.259 Million cell updates/sec

Title: US-09-800-065-1
Perfect score: 1920
Sequence: 1 atggcgaatttctctaa.....cgttgagctaaacaatatg 1920

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 24791104 seqs, 12571243825 residues
Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US099D_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US100A_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US101A_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US101B_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US102A_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US102B_COMB.seq.*

44: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US6023_COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
75: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
76: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
77: /cgn2_6/ptodata/2/pna/US6033_COMB.seq.*
78: /cgn2_6/ptodata/2/pna/US6034_COMB.seq.*
79: /cgn2_6/ptodata/2/pna/US6035_COMB.seq.*
80: /cgn2_6/ptodata/2/pna/US6036_COMB.seq.*
81: /cgn2_6/ptodata/2/pna/US6037_COMB.seq.*
82: /cgn2_6/ptodata/2/pna/US6038_COMB.seq.*
83: /cgn2_6/ptodata/2/pna/US6039_COMB.seq.*
84: /cgn2_6/ptodata/2/pna/US6040_COMB.seq.*
85: /cgn2_6/ptodata/2/pna/US6041_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	1920	100.0	1920	31	US-09-800-065-1
2	1920	100.0	5513	31	US-09-800-065-3
3	597.6	31.1	7088	4	US-08-029-329-1
4	91.2	4.8	7814	66	US-60-226-176-1986
5	91.2	4.8	7814	67	US-60-233-468-1986
6	91.2	4.8	7814	75	US-60-313-371-1986
c 7	76.8	4.0	7814	66	US-60-226-176-1986
c 8	76.8	4.0	7814	67	US-60-233-468-1986
c 9	76.8	4.0	7814	75	US-60-313-371-1986
c 10	74.2	3.9	540	1	PCT-US00-41008-1
c 11	74.2	3.9	186935	66	US-60-226-176-1976
c 12	74.2	3.9	186935	67	US-60-233-468-1976
c 13	74.2	3.9	186935	75	US-60-313-371-1976
c 14	71.2	3.7	961	80	US-60-360-207-31796
c 15	69.6	3.6	186935	66	US-60-226-176-1976
c 16	69.6	3.6	186935	67	US-60-233-468-1976
c 17	69.6	3.6	186935	75	US-60-313-371-1976
c 18	63.8	3.3	486	1	PCT-US00-22496-13
c 19	63.8	3.3	486	24	US-09-639-207-13
c 20	60.4	3.1	7218	8	US-08-466-194-14
c 21	59.2	3.1	510	1	PCT-US01-00563-15232

```
c 22 59.2 3.1 510 33 US-09-864-761-18737 Sequence 18737, A
c 23 59.2 3.1 510 41 US-10-182-993-14763 Sequence 14763, A
c 24 59.2 3.1 510 41 US-10-182-995-11883 Sequence 11883, A
c 25 59.2 3.1 510 42 US-10-203-134-15069 Sequence 15069, A
c 26 59.2 3.1 510 42 US-10-203-136-15065 Sequence 15065, A
c 27 59.2 3.1 510 42 US-10-203-137-15232 Sequence 15232, A
c 28 59.2 3.1 510 42 US-10-203-138-7146 Sequence 7146, Ap
c 29 59.2 3.1 510 42 US-10-203-139-14660 Sequence 14660, A
c 30 59.2 3.1 510 67 US-60-236-359-9553 Sequence 9553, Ap
c 31 57.6 3.0 604 64 US-60-207-458-24992 Sequence 24992, A
c 32 57.6 3.0 3666 15 US-09-137-531-13 Sequence 13, Appl
c 33 57.6 3.0 3666 15 US-09-137-531-14 Sequence 14, Appl
c 34 57.6 3.0 4197 15 US-09-137-531-7 Sequence 7, Appl
c 35 57.6 3.0 4197 15 US-09-137-531-8 Sequence 8, Appl
c 36 57.4 3.0 390 21 US-09-540-234-2089 Sequence 2089, Ap
c 37 56.8 3.0 961 80 US-60-360-207-31796 Sequence 31796, A
c 38 56.8 3.0 1954 80 US-60-360-207-4707 Sequence 4707, Ap
c 39 56.8 3.0 8307 63 US-60-258-275-230 Sequence 230, App
c 40 56 2.9 632 33 US-09-865-439A-30490 Sequence 30490, A
c 41 56 2.9 632 64 US-60-207-458-74787 Sequence 74787, A
c 42 56 2.9 897 80 US-60-360-207-39592 Sequence 39592, A
c 43 55.8 2.9 1118 80 US-60-360-207-33054 Sequence 33054, A
c 44 55.4 2.9 1236 39 US-10-077-584-3 Sequence 3, Appl
c 45 55.4 2.9 1236 70 US-60-269-157-3 Sequence 3, Appl
```

ALIGNMENTS

RESULT 1

US-09-800-065-1

; Sequence 1, Application US/098000065

; GENERAL INFORMATION:

; APPLICANT: Bouyer, Donald H

; APPLICANT: Croquet-Valdes, Patricia

; APPLICANT: Stenos, John

; APPLICANT: Walker, David H

; TITLE OF INVENTION: Rickettsia felis Outer Membrane Protein

; FILE REFERENCE: 026.00121

; CURRENT APPLICATION NUMBER: US/09/800,065

; CURRENT FILING DATE: 2001-03-06

; PRIOR FILING DATE: 2000-03-06

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1920

; TYPE: DNA

; ORGANISM: Rickettsia felis

US-09-800-065-1

```
Query Match 100.0%; Score 1920; DB 31; Length 1920;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ATGCGCAATATTTCTTAAATATTTTCAAAAGCAATTTCAAAAGCTCTTAAACCTGCT 60
DB 1 ATGCGCAATATTTCTTAAATATTTTCAAAAGCAATTTCAAAAGCTCTTAAACCTGCT 60
QY 61 TTATTCACCACTCAACCGGAGGATATGCTAACCGGTAGTGAGGCTTGGTGCTGCA 120
DB 61 TTATTCACCACTCAACCGGAGGATATGCTAACCGGTAGTGAGGCTTGGTGCTGCA 120
QY 121 AGAACCTTAAGTCTGATGTGAGAGCTTGCAGCGGAGCAATATAGTCTCTGGAGCC 180
DB 121 AGAACCTTAAGTCTGATGTGAGAGCTTGCAGCGGAGCAATATAGTCTCTGGAGCC 180
QY 181 GGTGCTTTTGTAGCGGTTCTACTTTACAATATACCGGTCTTTACGGTTACTGATGCT 240
DB 181 GGTGCTTTTGTAGCGGTTCTACTTTACAATATACCGGTCTTTACGGTTACTGATGCT 240
QY 241 GAGTAAGTCTCGTGCATTAGATTTAAATTTTTCAGCAGCTCTTTTTCAGTAAC 300
DB 241 GAGTAAGTCTCGTGCATTAGATTTAAATTTTTCAGCAGCTCTTTTTCAGTAAC 300
```

```
Db 241 GACGTAAGTGTTCGTGCATTAGATTTAAATTAATTTTTCAGCAGGCTCTTTTTCAGTAAC 300
QY 301 GGTGATATTTTCATAGGTTTCAGTGGTAGATACGGGAGAGCTAATAAATTTGCGAGTTAA 360
Db 301 GGTGATATTTTCATAGGTTTCAGTGGTAGATACGGGAGAGCTAATAAATTTGCGAGTTAA 360
QY 361 ATTGATGATGTTTAAACCTTAACTTTAACAGGTACCGGTACTGAGCCTACGGTGCAAA 420
Db 361 ATTGATGATGTTTAAACCTTAACTTTAACAGGTACCGGTACTGAGCCTACGGTGCAAA 420
QY 421 CCTGCTGTTTATTTCCAAAGTGGACAGCTGCTGCTAATAATACATATACATGCTTTAG 480
Db 421 CCTGCTGTTTATTTCCAAAGTGGACAGCTGCTGCTAATAATACATATACATGCTTTAG 480
QY 481 AATAAATCTCTAGTGGAGCAATGCGGTTTGAATTTGCTTCAGATCCAGATGATTA 540
Db 481 AATAAATCTCTAGTGGAGCAATGCGGTTTGAATTTGCTTCAGATCCAGATGATTA 540
QY 541 GGACCAATAACGCTTTCAGGAGGAGTATAGATGGAGGAGTATATAAATGACAATACAG 600
Db 541 GGACCAATAACGCTTTCAGGAGGAGTATAGATGGAGGAGTATATAAATGACAATACAG 600
QY 601 GCTGCCATTACGGAACAATAGGTAACTAATCCGAGCTCAAAATAGCATTTGGAGCA 660
Db 601 GCTGCCATTACGGAACAATAGGTAACTAATCCGAGCTCAAAATAGCATTTGGAGCA 660
QY 661 AGTACGCTTTCTCTTGGAGGCGCAGTTATTAAGCCACTACGACTAAGTTAAACAAATG 720
Db 661 AGTACGCTTTCTCTTGGAGGCGCAGTTATTAAGCCACTACGACTAAGTTAAACAAATG 720
QY 721 GCGCGGTATTAACACCTTACAAATGCAAAATGCAATTAAGTAAACAGTGTGTTGATA 780
Db 721 GCGCGGTATTAACACCTTACAAATGCAAAATGCAATTAAGTAAACAGTGTGTTGATA 780
QY 781 ACAGCGGTGATGATGTAGTGTCTTAAATTTAAACGAGGCTTGAAGTCAAGTAACCTG 840
Db 781 ACAGCGGTGATGATGTAGTGTCTTAAATTTAAACGAGGCTTGAAGTCAAGTAACCTG 840
QY 841 AATATAGGTATTAACAAATTCATTAGCGACAATAAGTATAGGAGCAGGTACGGTACGTT 900
Db 841 AATATAGGTATTAACAAATTCATTAGCGACAATAAGTATAGGAGCAGGTACGGTACGTT 900
QY 901 GGGGAGCGGTTTAAAGCCACTACGACTAAGTTACGAATGAGGCTCGGTATTAACA 960
Db 901 GGGGAGCGGTTTAAAGCCACTACGACTAAGTTACGAATGAGGCTCGGTATTAACA 960
QY 961 CTTACAAATGCAGTATTAACAGGTGCTGTGATAAACACCAGCGGTGATAATGTAGT 1020
Db 961 CTTACAAATGCAGTATTAACAGGTGCTGTGATAAACACCAGCGGTGATAATGTAGT 1020
QY 1021 GTCGTAATCTAAGTGGAGCATTGAGTCAAGTAACCGGGAATATAGTAATAACAAATTC 1080
Db 1021 GTCGTAATCTAAGTGGAGCATTGAGTCAAGTAACCGGGAATATAGTAATAACAAATTC 1080
QY 1081 TTAGCCACGATAATATAGGAGCAGGTGTGCTACCTTTGGATGGAGCGGTTATTAAGCT 1140
Db 1081 TTAGCCACGATAATATAGGAGCAGGTGTGCTACCTTTGGATGGAGCGGTTATTAAGCT 1140
QY 1141 ACTACGACTAAGTTAACAGATGATGCTCAGTATTAATATTTACGAATCTCTGTAGTAG 1200
Db 1141 ACTACGACTAAGTTAACAGATGATGCTCAGTATTAATATTTACGAATCTCTGTAGTAG 1200
QY 1201 ACCGGAGCAATAGATATACCGGTAAATGCAATTAAGGTGTGTTAATCTTTACCGGAG 1260
Db 1201 ACCGGAGCAATAGATATACCGGTAAATGCAATTAAGGTGTGTTAATCTTTACCGGAG 1260
QY 1261 AGTACGTAACCGATATATAGGTACACGCGAGTATTACGAGGTAAAGCTTAGGAGCA 1320
Db 1261 AGTACGTAACCGATATATAGGTACACGCGAGTATTACGAGGTAAAGCTTAGGAGCA 1320
QY 1321 GGTTCGTCGCAATACAAAGCGGAGTAGTAAAGCGCAATGCAATTAACCTTTAAACGGATA 1380
Db 1321 GGTTCGTCGCAATACAAAGCGGAGTAGTAAAGCGCAATGCAATTAACCTTTAAACGGATA 1380
```

QY	1381	GGTCTAGTAGTAACATTTTACCGGTGATAGTACGGTAAACAGTACTAGTAGGTGGTACAGAA	1440
Db	1381	GGTCTAGTAGTAACATTTTACCGGTGATAGTACGGTAAACAGTACTAGTAGGTGGTACAGAA	1440
QY	1441	CCTTCGCAACAGTGAATATAGGACGACGAATAACATTTACAGCGCGGAGGAAAGCCTAGCT	1500
Db	1441	CCTTCGCAACAGTGAATATAGGACGACGAATAACATTTACAGCGCGGAGGAAAGCCTAGCT	1500
QY	1501	GCGAATAATATAGATTTTGGAGCTCGAGTAGTAATTTAGAGTTTAAACGGTCTCTCGCGGTAAAG	1560
Db	1501	GCGAATAATATAGATTTTGGAGCTCGAGTAGTAATTTAGAGTTTAAACGGTCTCTCGCGGTAAAG	1560
QY	1561	AATTATACTTAATCGGAAGTATAGCAACCGGTAAATAGTACACATTAATATTAATGCT	1620
Db	1561	AATTATACTTAATCGGAAGTATAGCAACCGGTAAATAGTACACATTAATATTAATGCT	1620
QY	1621	GCTGGTACAGTGAATGTCRAATGATCTTAGTAGTAGCACAGTTGCACAAATTAACATTCAA	1680
Db	1621	GCTGGTACAGTGAATGTCRAATGATCTTAGTAGTAGCACAGTTGCACAAATTAACATTCAA	1680
QY	1681	AATAATAAGATTTTCTTAATAATGCTAAGAACGCTGATGTTGATATATTAGACGCTCAG	1740
Db	1681	AATAATAAGATTTTCTTAATAATGCTAAGAACGCTGATGTTGATATATTAGACGCTCAG	1740
QY	1741	GCAGTAGTTTAAAGGACGACGCTCAGCTCTTTTCTTAGCTACAGTTAGTCTACAGATG	1800
Db	1741	GCAGTAGTTTAAAGGACGACGCTCAGCTCTTTTCTTAGCTACAGTTAGTCTACAGATG	1800
QY	1801	ATAGAGTTATCACTCTTAAATAATCATTTACCGGTCCTTGGCTAACCGGTGGTGGTAA	1860
Db	1801	ATAGAGTTATCACTCTTAAATAATCATTTACCGGTCCTTGGCTAACCGGTGGTGGTAA	1860
QY	1861	TTCTTTTGGAGTCCCAACAAAACCTTATGACCGTACAAAGGTGACGTTGGAGCTAAAAAATAG	1920
Db	1861	TTCTTTTGGAGTCCCAACAAAACCTTATGACCGTACAAAGGTGACGTTGGAGCTAAAAAATAG	1920

RESULT 2

US-09-800-065-3

; Sequence 3, Application US/098000065

; GENERAL INFORMATION:

; APPLICANT: Bouyer, Donald H

; APPLICANT: Crocquet-Valdes, Patricia

; APPLICANT: Stenos, John

; APPLICANT: Walker, David H

; TITLE OF INVENTION: Rickettsia felis Outer Membrane Protein

; FILE REFERENCE: 026.00121

; CURRENT APPLICATION NUMBER: US/09/800,065

; CURRENT FILING DATE: 2001-03-06

; PRIOR APPLICATION NUMBER: US 60/187,323

; PRIOR FILING DATE: 2000-03-06

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 5513

; TYPE: DNA

; ORGANISM: Rickettsia felis

US-09-800-065-3

Query Match		100.0%;	Score 1920;	DB 31;	Length 5513;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches	1920;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	
QY	1	ATGCGGAATATTTCTCTAAATATTTTCAAAAGCAATTTCAAAAGCTCTTAAACTGCT	60		
Db	478	ATGCGGAATATTTCTCTAAATATTTTCAAAAGCAATTTCAAAAGCTCTTAAACTGCT	537		
QY	61	TTATTACCACCTCAACCGCAGGATAATGCTAAACGGTAGTGGAGTCTTTGGTCTGCA	120		
Db	538	TTATTACCACCTCAACCGCAGGATAATGCTAAACGGTAGTGGAGTCTTTGGTCTGCA	597		
QY	121	AGAACCGTAACTGCTGATGGTGACAGCTTGCAGCGGCAACAAATATAGGTCCTGGAGCC	180		

Db 1678 ACCGGAGCAATAGATAATACCGGTAAATGCCAATAAAGGTGGTAACTTTTACCAGGAGCA 1737
Qy 1261 AGTACGGTAACCGATAATATAGTTAACCGCAGTATTAGCAGAGGTAAAGCGTAGGAGCA 1320
Db 1738 AGTACGGTAACCGATAATATAGTTAACCGCAGTATTAGCAGAGGTAAAGCGTAGGAGCA 1797
Qy 1321 GGTTCGCTGCAATACACAGCGGAGTAGTAAAGCGAATGCAATAAACTTACGGATAT 1380
Db 1798 GGTTCGCTGCAATACACAGCGGAGTAGTAAAGCGAATGCAATAAACTTACGGATAT 1857
Qy 1381 GCGTCAGTAGTAACATTTACCGGTGATAGTACGGTAACAGGTAGTATAGGTGGTACAGAA 1440
Db 1858 GCGTCAGTAGTACATTTACCGGTGATAGTACGGTAACAGGTAGTATAGGTGGTACAGAA 1917
Qy 1441 CTCCTCGCAACAGTGAATATAGGAGCAGGAATAACATTTACGAGCGGAGGAGCCTAGCT 1500
Db 1918 CTCCTCGCAACAGTGAATATAGGAGCAGGAATAACATTTACGAGCGGAGGAGCCTAGCT 1977
Qy 1501 GCGAATAATATAGATTTTGGAGCTCGGAGTAATTTAGAGTTTACGGTCTTGGCGGTAA 1560
Db 1978 GCGAATAATATAGATTTTGGAGCTCGGAGTAATTTAGAGTTTACGGTCTTGGCGGTAA 2037
Qy 1561 AATTATAACTTAATCGGAACCTATAGCAACCGTAAATATGCTACACCTTAATATTAATGCT 1620
Db 2038 AATTATAACTTAATCGGAACCTATAGCAACCGTAAATATGCTACACCTTAATATTAATGCT 2097
Qy 1621 GCTGTCAGTATGCAATATGATTTAGTAGTACGAGCAGTTGCAACAAATTAACATTCAA 1680
Db 2098 GCTGTCAGTATGCAATATGATTTAGTAGTACGAGCAGTTGCAACAAATTAACATTCAA 2157
Qy 1681 AATAATAGATTTTGTAAATAGCTAAGAACGCTGATTTGATATATATAGAGCCTCAG 1740
Db 2158 AATAATAGATTTTGTAAATAGCTAAGAACGCTGATTTGATATATATAGAGCCTCAG 2217
Qy 1741 GCGATTTAGTTTAAAGGAGCAGTTCAGCTCTTTCTTAGTAACTAGTTAGTCTACAGATG 1800
Db 2218 GCGATTTAGTTTAAAGGAGCAGTTCAGCTCTTTCTTAGTAACTAGTTAGTCTACAGATG 2277
Qy 1801 ATAGATTTATCAGCTTTAAATAATCATTTACCGGTCTTTGCTAAACGGTGGTGGTGAAT 1860
Db 2278 ATAGATTTATCAGCTTTAAATAATCATTTACCGGTCTTTGCTAAACGGTGGTGGTGAAT 2337
Qy 1861 TTCTTTTGGTCCCAACAACTTATGACCGTACAGGTGAGGTGGAGCTAAACAAATAG 1920
Db 2338 TTCTTTTGGTCCCAACAACTTATGACCGTACAGGTGAGGTGGAGCTAAACAAATAG 2397

RESULT 3

US-08-029-329-1
; Sequence 1, Application US/08029329
; GENERAL INFORMATION:
; APPLICANT: SUMNER W., JOHN
; APPLICANT: ANDERSON E., BURT
; APPLICANT: PELLET E., PHILIP
; APPLICANT: SANCHEZ-MARTINEZ, DEMETRIO
; TITLE OF INVENTION: "BACULOVIRUS EXPRESSED 190 kDa ANTIGENS
; TITLE OF INVENTION: OF RICKETTSIA RICKETTSII AND RELATED METHODS"
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/029,329
; FILING DATE: 19930309

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414,049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7088 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 70..6816
; US-08-029-329-1

Query Match 31.1%; Score 597.6; DB 4; Length 7088;
Best Local Similarity 63.9%; Pred. No. 7.1e-143;
Matches 982; Conservative 0; Mismatches 518; Indels 36; Gaps 4;
Qy 1 ATGGCGAATATTTCTTAAATATTTCAAAAAGCAATTTCAAAAGGTCTTTAAACCTGCT 60
Db 70 ATGGCGAATATTTCTTAAATATTTCAAAAAGCAATTTCAAAAGGTCTTTAAACCTGCT 129
Qy 61 TTATTACCACCTCAACCGCAGCATATGCTAACCGGTAGTGGAGTCTTGGTGGTCA 120
Db 130 TTATTACCACCTCAACCGCAGCATATGCTAACCGGTAGTGGAGTCTTGGTGGTCA 189
Qy 121 AGAACCTTAATGCTGATGCTGAGAGCTTGCAGCGGCAACAAATATAGTCTTGGAGCC 180
Db 190 ACAGGTGTTATGCTACTAATAATATGACGACATTTAGTAAATATGTTGCAATAAAT 249
Qy 181 GGTGCTTTTGTAGCGGGTCTTACTTTACATATACCGGTGCTTTTACGGTTCGTGCT 240
Db 250 TGGAAATGAGATAACGGTGCAGGGGTAGCTAATGGTACTCTGCTGCGGTCTTCAAAAC 309
Qy 241 GACGTAAGTGTTCGTGCTATGATTTAAATAATTTTCACGACGCTCTTTTTCAGTAAT 300
Db 310 AATTGGCATTTACTTACGGTGGTATTTACTGCTACGATGACGATGACGGGATCGTATT 369
Qy 301 GGTGATATTTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACT 360
Db 370 ATTAAGCTATATAATGTTGCGGTACTTACTTACTTACTTACTTACTTACTTACTTACT 429
Qy 361 ATGATG-----ATGTTTAACTTAACTTTAACTTTAACTTTAACTTTAACTTTAACT 399
Db 430 GTCGTTGGTTCGATTTAAGCAAGGTAACTTTGCTGCTTACTTCTTAATGCGGCAAA 489
Qy 400 ACTGACGCTACGGTGCATAATCTCGCTGTTTATTTCAAGGTGCAGAGCTGCTGCTAAT 459
Db 490 AGCTTAATTAATGTTAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 549
Qy 460 AATACATATATCTGTTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACT 519
Db 550 GATAATTTATACAGTTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACT 608
Qy 520 GCTTCAGATCCAGATGTTATGAGCAACCAATTAACGTTGACAGGAATATAGATGAGGAGGT 579
Db 609 -----ACAATCTCGAGTCCGTCAGAAAGTAACTTTGCAAGAAATATAGATGAGGAGGT 663
Qy 580 ATAATACTGACAAATACAGATGCTGCAATTTAAGCAACAAATAGTAACTTAACTTAACT 639
Db 664 ATAATACTGACAAATACAGATGCTGCAATTTAAGCAACAAATAGTAACTTAACTTAACT 723
Qy 640 GCTCAATAAGCATTTGAGCAAGTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 699
Db 724 GCAACAGTGAATGTAGGAGCAGGTACAGCCACGTTAGGGGAGCGGTATTTAAAGCTACT 763
Qy 700 ACAGCTAAGTTAACAATGCAGCGCGCGGTATTTAACTTAACTTAACTTAACTTAACT 759


```
Db 2988 TTGCTGTTGCTGTTGCTGATGACTGTTGCTATCGTTGCTGTTTGGTTATGTTGCTGCTGTTGTTG 3047
QY 1076 ATTCATTAGCCOACGATAAATATAGGACGAGGTGGTGTACCTTGGATGGAGCGGTATTA 1135
Db 3048 CTGCTGCTGCTACTGTTGTTGTTGCTGCTGCTCTATTTGTTGCTGCTGCTGTTGCTG 3107
QY 1136 AAGCTACTACGACACAAAGTTAAAC---AGATGATGCGTACGATTAATATTTACGAATCCTG 1192
Db 3108 TTGTTACTGCTGCTATTTGTTGCTGCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3167
QY 1193 TAGTAGTAACCGGAGCAATAGATAATACCGGTAATGCCAATAAGGTTGGTAATCTTTA 1252
Db 3168 TTGCTGCTACTATTTACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3227
QY 1253 CCGGAGCAAGTACGCTAACCGGATAATATAGGTAACAGCGGAGTATTAGCAGAGGTAAGCG 1312
Db 3228 TTGTTGCTACTATGTAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3287
QY 1313 TAGGAGCAGGTTTCTGCAATATACAGCGGAGTAGTAAAGCGAATGCAATAAATCTTAA 1372
Db 3288 TTGCTACTGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3347
QY 1373 CGGATAATGCTGCTAGTAGTAACATTTACCGGTGATAGTACGGTAACAGGTAATAGGTG 1432
Db 3348 TTGCTATTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3407
QY 1433 GTACAGAACTCTTTCGCAACAGTGAATATAGGACGAGGAATAACATTTACGAGCCGAGGAA 1492
Db 3408 TTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3467
QY 1493 GCCTAGCTGCGAATAATAGATTTTGGAGCTGCGAGTAATTTAGAGTTTAAACGGTCTG 1552
Db 3468 CTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3527
QY 1553 CCGGTAAGAATTTAACTTAATCGGAATATAGCAACGGTAAATGCTACACTTAATA 1612
Db 3528 CTGTTGCTGTTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3587
QY 1613 TTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1672
Db 3588 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3647
QY 1673 ACATTCAAAATAAAGATTTTGTAAATAATGCTAAGAACGCTGATGTTGATATATTAG 1732
Db 3648 TTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3707
QY 1733 ACGCTACGGCGATGATGTTTT 1752
Db 3708 CTACTATTGCTGCTGCTGTT 3727
```

RESULT 7

```
US-60-226-176-1986/c
; Sequence 1986, Application US/60226176
; GENERAL INFORMATION:
; APPLICANT: Ring, Huijun Z.
; APPLICANT: Malsen, Gareth
; APPLICANT: Townley, David
; APPLICANT: Morris, Macdonald
; TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated with ADME Genes
; FILE REFERENCE: GX-0013-1 P
; CURRENT APPLICATION NUMBER: US/60/226,176
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2447
; SOFTWARE: PERL Program
; SEQ ID NO 1986
; LENGTH: 7814
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GBI:AC024253_3_000004
US-60-226-176-1986
```

```
Query Match 4.0%; Score 76.8; DB 66; Length 7814;
Best Local Similarity 43.0%; Pred. No. 2.2e-08;
Matches 433; Conservative 0; Mismatches 572; Indels 3; Gaps 1;

QY 509 GTTTGACTATTTGCTTCAGATCCAGATGTATTAGGACCAATAACGCTTGCAGGAATATAG 568
Db 3851 GATTAAACAGCAGCATCAATAGCCACAGCAACAGCAACAGCAGCAGTACGACAGAAATCG 3792
QY 569 ATGAGAGAGGTATTAATTAAGTGACAAATACAGATGCTGCCATTTACGGACAAATAGGTATA 628
Db 3791 CAGTAATAACAGTAGCAGCAGCAGCAACATTCATTACCATAGTAGCAACAACAGTAG 3732
QY 629 CTAATCCGGCAGCTCAATAAGCAATTGGAGCAAGTAGCGTTTCTTGGAGGGCGCTTA 688
Db 3731 CAGCAACACAGCAGCAATAGTAGCAGCAATAGCAACAGTAATAGTAGCAGCAGCAACAA 3672
QY 689 TTAAGCCACTACGACTAAGTTTAAACAAATGCGAGCGCGGTATTAACTTACAAATGCAA 748
Db 3671 TAGCAACAGTAGCAGCTAGCAACAAACAGCAACAGCAGCAGCAACAAATAGAGCAGCAG 3612
QY 749 ATGCAAGTATTAAACAGGTGCTGTTGATTAACACCAGCGGGGTGATGTAGGTGCTCTAA 808
Db 3611 CAGCAACAACAGTAGTAGCAGCAGCAACAACAGCAACAATAACAAAAGCAACAGTAG 3552
QY 809 ATTTAAACGGAGCGTTGAGTCAAGTCAAGTAACTGGAAATATAGGTAAATACAAATTCATTAGCGA 868
Db 3551 CAACAGTATCAGCAACAGCAACAGCATCAGCAACAACAGCAACATAGCAGCAGTAAAAA 3492
QY 869 CAATAAGTGTAGGAGCAGGTACGGTACGTTAGGGGGAGCGGTTATTAAAGCCACTACGA 928
Db 3491 CAGCAACAGCAGCAGCAACAAACAGCAACAGCAGCAGCAACAATAAGAGCAGCAACACAG 3432
QY 929 CTAAGTTAAAGATGACGCGTGGTATTAACTTACAAATGCAAGTATTAAACAGGTGCTG 988
Db 3431 CAACAGTGTAGCAGCAGCAACAACAGCAGCAACAACAGCAACAATAAGCAGCAGTAT 3372
QY 989 TTGATAACACACAGCGCGGTGATAAATGATAGGTGCTGTAATCTAAGTGGAGC---ATTGA 1045
Db 3371 CAGCAACAGCAACAGCAATAGCAACAAAGCATCAGCAACAGCAATAGCATCACAATTA 3312
QY 1046 GTCAAGTAAACCGGGAATATAGGTAAATACAAATTCATTAGCCAGCAATAATAGGACGAG 1105
Db 3311 CAACAGCAACAGCAGCAGTAGTAGCAACAGAAATCAGAGTAACAGCAACAGTAGCAGCAGCAG 3252
QY 1106 GTGTGCTACCTTGGATGGAGCGGTTATTAAAGCTACTAGAGCTTAAGTTAACAGATGATG 1165
Db 3251 CAATAATTACCATAGTAGCAACAACAGTAGCAGCAACACCAGCAGCAGTAGCAGCAGCAA 3192
QY 1166 GTTCAGTATTAAATTTACCAATCCTGTAGTAGTAACCGGAGCAATAGATAATACCGGTA 1225
Db 3191 TAGCAACAGTAATAGTAGTAGCAACAACAGTAATAGCAACAGTAGCAGCATCAGCAACAACA 3132
QY 1226 ATGCCAATAAAGGTGTGGTAACTCTTTACCGGAGCAAGTACGGTAACCGAATAATATAGTA 1285
Db 3131 GAGCAACAATAGCAGCAGTAGTAACAACAGCAACAGCAGCAACAACAGCAACAATAAGAGCAGCAG 3072
QY 1286 ACACGGCAGTATTAGCAGAGGTAAAGCGTAGGAGCAGGTTTGTGCAAAATACAAAGCGGAG 1345
Db 3071 CAACAACACAGTAGCAGCAGCAGCAACAACAGCAACAATAACAAAAGCAACGATAGCAAA 3012
QY 1346 TAGTAAAGCGAATGCAATAAATTAACGGATTAACGGATTAACGGTACGTAGTAAATTTACCGGTG 1405
Db 3011 CAGTATCAGCAACAGCAACAGCAATAGCAACAGCAGCAGCATCAGCAACAATAAGCAGCAG 2952
QY 1406 ATAGTAGCGTAAACAGGTAGTATAGTGGTTCGGAACACTCTTCGGAACAGTCAATATAGGAG 1465
Db 2951 TAAACAGCAACAGCAGCAGCAACAACAGCAACAACAGCAGCAGCAACAATAAGAGCAGCAG 2892
QY 1466 CAGGAATAACATTTACGAGCGCGGAGGAGCGCTAGCTGCGAATAATATAG 1513
Db 2891 CAACAGCAACAGTAGCAGCAGCAGCAACAACAGCAACAATAAGCAACAATAAGCAACAAG 2844
```

```

RESULT 8
US-60-233-468-1986/c
; Sequence 1986, Application US/60233468
; GENERAL INFORMATION:
; APPLICANT: Ring, Huijun Z.
; APPLICANT: Malsen, Gareth
; APPLICANT: Townley, David
; APPLICANT: Morris, MacDonald
; APPLICANT: Valdes, Ana
; TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated with ADME Genes
; FILE REFERENCE: GX-0013-2 P
; CURRENT APPLICATION NUMBER: US/60/233,468
; CURRENT FILING DATE: 2000-09-18
; NUMBER OF SEQ ID NOS: 2488
; SOFTWARE: PERL Program
; SEQ ID NO 1986
; LENGTH: 7814
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GBI:AC024253_3_000004
US-60-233-468-1986

```

Query Match	4.0%;	Score 76.8;	DB 67;	Length 7814;
Best Local Similarity	43.0%;	Pred. No. 2.2e-08;		
Matches 433;	Conservative 0;	Mismatches 572;	Indels 3;	Gaps 1;
QY	509	GTTTGACTATTGCTTCAGATCCAGATGTTATAGGACCAATAACGCTTGCAGGAATAATAG	568	
Db	3851	GATTACAGCAGCATCAATAGCCACAGCACAGCAACAGCAGCAGTAGTACCAACAGAAATCG	3792	
QY	569	ATGGAGGAGGTATTAATAACTGTGCAATACAGATGCTGCCATTAACGGGAACAATAGGTAAATA	628	
Db	3791	CAGTAATAACAGTAGCAGCAGCAGCAGCAATCATTCATAGTAGCAACAACACAGTAG	3732	
QY	629	CTAATCCGGCAGTCAATTAAGCATTCGAGCAAGTAGCGCTTCTCTTGGGGGCGCAGTTA	688	
Db	3731	CAGCAACACCAGCAGCAGTATAGTAGCAGCAATAGCAACAGTAATAGTAGCAGCAACAGCA	3672	
QY	689	TTAAGCCACTACGACTTAAGTTTAACAAATCAGCGCCGGTATTAAACACTTTACAATGCAA	748	
Db	3671	TAGCAACAGTAGCAGCATCAGCAACAACAGCAGCAGCAACAATAAGACGACGAC	3612	
QY	749	ATCCAGTATTAAACAGGTGCTTGTGATTAACACCACAGCAGCGGTGATGTAGGTGTCATTAA	808	
Db	3611	CAGCAACAACAGTAGTAGCAGCAGCAGCAACAACAGCAACAATAACAANAAGCAACGATAG	3552	
QY	809	ATTTAACGGGAGCGTTCAGTCAAGTCAAGTACTGGAATATAGGTAATACAAATTCATTAGCGA	868	
Db	3551	CAACAGTATCAGCAACAGCAACAGCATCAGCAACAACAGCAACAATAGCAGCAGTAAAA	3492	
QY	869	CAATAAGTGTAGNAGCAGGTACGGCTAGTGTAGGGGAGCGGTTTATTAAAGCCACTACGA	928	
Db	3491	CAGCAACAGCAGCAGCAACAACAGCAACAGCAGCAGCAACAATAAGAGCAGCAACAACAG	3432	
QY	929	CTAAGTTTAAAGCAATGCAAGCGTTCGGTATTAAACACTTACAAATGCAGTATTAAACAGGTCGT	988	
Db	3431	CAACAGTAGCAGCAGCAGCAACAACAGCAGCAACAACAAGCAACAATAGCACAGTAT	3372	
QY	989	TTGATTAACACACAGCGCGGTGATAATGTAGGTGTGCTTAATCTAAGTGGAGC---ATTGA	1045	
Db	3371	CAGCAACAGCAACAGCAATAGCAACAACAAGCATCAGCAACAGCAATAAGCATCACAATTA	3312	
QY	1046	GTCAAGTAAACCGGGAATATAGGTAATACAAATTCATTAGCCAGATAAATATAGGACGAG	1105	
Db	3311	CAACAGCAACAGCAGCAGTAGCAACAGAAATCAGTAAACAAGTAGCAGCAGCAGCAG	3252	
QY	1106	GTGTGGCTACCTTGGATGGAGCGGTTATTAAAGCTACTACGACTTAAGTTAACAGATGATG	1165	
Db	3251	CAATAATTATCCATAGTAGCAACAACAGTAGCAACAACAGCAGCAGTATAGCAGCAGCA	3192	

```

RESULT 9
US-60-313-371-1986/c
; Sequence 1986, Application US/60313371
; GENERAL INFORMATION:
; APPLICANT: Ring, Huijun Z.
; APPLICANT: Malsen, Gareth
; APPLICANT: Townley, David
; APPLICANT: Morris, MacDonald
; TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated with ADME Genes
; FILE REFERENCE: GX-0013-5 P
; CURRENT APPLICATION NUMBER: US/60/313, 371
; CURRENT FILING DATE: 2001-08-16
; NUMBER OF SEQ ID NOS: 2447
; SOFTWARE: PERL Program
; SEQ ID NO 1986
; LENGTH: 7814
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GBI:AC024253_3_000004
US-60-313-371-1986

```

	Query Match	4.0%; Score 76.8; DB 75;	Length 7814;
	Best Local Similarity	43.0%; Pred. No. 2.2e-08;	
Matches	Conservative	0;	Mismatches 572; Indels 3; Gaps 1;
QY	509 GTTTGACTATTGCTTCAGATCCAGATGTATTAGGCCAAATAACGCCTTGCAGGAATAATAG	568	
Dd	3851 GATTAAACAGCAGCATCAATAGCCACACAGACAACAGCAGTAGTGACCAAGAATCG	3792	
QY	569 ATGGAGGAGGTATAAATAACTTGACAATACAGATGCTGCCATTTAACGGGAACACATAGTGA	628	
Dd	3791 CAGTAATAACAGTAGCAGCAGCAGCACAGCAATCATTTACCATAGTAGCAACACAGTAG	3732	
QY	629 CTAAATCCGGCAGCTCAAATAAGCATTTGGAGCAAGTAGCTTTCTCTTTGGAGGGGCAGTTA	688	
Dd	3731 CAGCAACACCAGCAGCAATAGTAGCAGCAANTAGCAACAGTAATAGTAGCAGCAACAGCAA	3672	
QY	689 TTAAGGCCACTACGACTAAGTTTAAACAATCAGGCGCCGGTATTAAACACTTTACAANTCGAA	748	
Dd	3671 TAGCAACAGTAGCAGCATCAGCAACAACAGCAACAGCAGCAGCAACAATAAGAGCAGCAG	3612	
QY	749 ATGCAGTATTAAcAGGTGCTGTTCATAAACCAACAGCGGGTGATGATGTAGGTGCTCTTAA	808	
Dd	3611 CAGCAACAACAGTAGTAGCAGCAGCAGCAACAGCAACAANTAAcAAAAGCAACAGTAGT	3552	
QY	809 ATTtAAACGGAGCGTTTCAGTCAAGTAACTCGGAATATAGGTAAATPAAAAATTCATTAGCGA	868	


```

Db      7612  TTGCTGCTGCTGCTGCTGTTGTTACTGCTGCTATTGTTGCTGTTATTGTTGCTGATGCTG  7553
QY      926   CGACTAAGTTAAACGAATGCAGCTCGGTATTAAACACTTACAAGTCAGTATTAAACAGGTG  985
Db      7552  CTACTGTTGCTATTGCTGTTGCTGCTACTATTACTGTTGCTATTGCTGCTGCTGCTACTGCTG  7493
QY      986   CTGTTGATAACACACACAGCGCGGTGATAATGTTAG  1018
Db      7492  CTGGTGTGCTGCTACTGTTGTTGCTACTATGG  7460

RESULT 13
US-60-313-371-1976/c
; Sequence 1976, Application US/60313371
; GENERAL INFORMATION:
; APPLICANT: Ring, Huijun Z.
; APPLICANT: Malsen, Gareth
; APPLICANT: Townley, David
; APPLICANT: Morris, MacDonald
; TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated with ADME Genes
; FILE REFERENCE: GX-0013-5 P
; CURRENT APPLICATION NUMBER: US/60/313,371
; CURRENT FILING DATE: 2001-08-16
; NUMBER OF SEQ ID NOS: 2447
; SOFTWARE: PERL Program
; SEQ ID NO 1976
; LENGTH: 186935
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GB:AC022322_3
US-60-313-371-1976

```

Query Match	3.98;	Score 74.2;	DB 75;	Length 186935;
Best Local Similarity	44.1%;	Pred. No. 3.4e-07;		
Matches 411;	Conservative 0;	Mismatches 513;	Indels 9;	Gaps 2;
QY	95	CCGCTAGTGGAGTCTCTGGTCTGCAAGAACCGTAACTGCTGATGGTCCAGAGCTTGCAG	154	
Db	8392	CTGCTACTGCTCGTGGTGTGCTGCTACTGCTGTGTGCTACTATGGTAATATTGCTGCTG	8333	
QY	155	CCGGAACAAATATAGGTCTCGAGCCGGTCTTTGTAGCCGGTCTACTTTACAATATA	214	
Db	8332	CTGCTGCTACTGTTGTACTGTGATTTCTGTGCTACTGCTGCTGCTGTTGTTAAT	8273	
QY	215	CCGCTGCTTTTACGGTFACTGATGCTGACGPAAGTGTTCGTGCATTAGATTTAAATAATT	274	
Db	8272	GTGATGTATTGCTGTGCTGATGCTTTGTGCTATTGCTGTGCTGCTGCTGCTGCTG	8213	
QY	275	TTGCAGCAGGTCCTTTTTCAGTAACCTGGTGATATTTCAATTAGGTTTCAGTGTAGATACGG	334	
Db	8212	TTGCTATTGTGCTTTGTTGCTGCTGCTGTTGTTGCTGCTGCTACTGCTGCTGCTG	8153	
QY	335	GAGGAGCTAATAACTTTCGACGTTAATATTGA--TGATGGTTTAAACCTTAACTTTAACAG	391	
Db	8152	TTGCTGCTCTATTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	8093	
QY	392	GTACCGCTACTGCAGCCTACGGTCAAACTCGCTGGTTGTTATCCAAAGTGGCAAGCTG	451	
Db	8092	CTGCTGCTATTGTTGCTGTTGTTGCTGATGCTGTGCTGCTGCTGCTGCTGCTGCTG	8033	
QY	452	CTGCTAATAATACATACTGCTTTAGSTAATAAATCTAGGTGGAGCGNAATGCCGGTT	511	
Db	8032	TTGCTTTTGTATTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	7973	
QY	512	TGACTATTGCTTCAGATCCAGATGTATTAGGACCAATAAGCGCTTCGAGGAATAATAGATG	571	
Db	7972	CTCTTATTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	7913	
QY	572	GAGGAGGTAAATAACTTGACAATACAGATGCTGCCATTACGAGNACAATAGGTAAATACTA	631	
Db	7912	TTGCTGCTACTATTACGTTGCTATTGCTGCTACTACTATTGCTGCTGCTGCTGCTGCTG	7853	

[illegible]

Query Match 3.7%; Score 71.2; DB 80; Length 961;
Best Local Similarity 46.4%; Pred. NO. 2.7e-07;
Matches 232; Conservative 0; Mismatches 268; Indels 0; Gaps 0;

[illegible]

471 TCCCTTACCTAATAATATACCTATGCTGGCGGAGATGCCATATGCTCAGATCC 530

Qy	876	TGTAGACGAGCTACGGTCACTTTAGGGGAGCGGTATTATTAAGCCACTAGCATAAGTT	935
Db	2326	AACAGCAGCGATAGCAGTGACAGCAGTGTATAGCAGTCACAGCACAAACAGCAGTGACAGT	2385
Qy	936	AACGAATGCAGCGTCCGGTATTAAACACTTTACAATTCACAGTGCAGTATTACAGCGTGTCTG	989
Db	2386	ACGGATAGCAGTGACAGCAGCAACAGCAGTGCACAGCAGTGATACAGTGCACAGCATGTAT	2445
Qy	990	-----TGATAACACCACAGCGGGTGATTAATGTAGTGTCGTAATATCTAAAGT-GAGCATTT	1043
Db	2446	AGTAGTGCAGCAGCAACAGCAGTGTATAGCAACGACAGCAGCAANTAGCAGTGACAGCAGT	2505
Qy	1044	GAGTCAAGTAACCGGGATATAGGTAATAACAATTCATTGCCACAGATAAATATAGCAGC	1103
Db	2506	GATAGCAGCAACAGCAGCATGATAGCAACACAGCAGTGTATGCAGTGTATGACAGTGACAGC	2565
Qy	1104	AGGTGTGGCTACTCTGGATGGAGCGGTTATTAAAGCTACTACGACTAAGTTAACAGATGA	1163
Db	2566	AGTGATAGCAGCAGCAGCAATAGCAGTCACAGCAGTAATACTAGTGACAGCAGCGATAGC	2625
Qy	1164	TGGTCAGTATTAAATTTACGAATCCCTGTGTAGTAGTAACCGGAGCAATAGATAATACCGG	1223
Db	2626	AACAGCAGTGATAGCAGCGCACAGCAGCATAGCAGTCACAGCAGTGATAGCAGCAGCAGC	2685
Qy	1224	TAATGCCAATAAAGGTGTGGTAACTTTTACCGGAGCAGTAGTCGTAACCGATATATAGG	1283
Db	2686	AATAGAAGTGACAGTAGTAATAGTAGTGACAGCAGCGATAGCAGTGACAGCAGCAACAGC	2745
Qy	1284	TAACCGGCAGCTATTAGCAGAGGTAAAGCGTAGCAGCAGGTTTGCTGCAAATACAAAGCGG	1343
Db	2746	AGTGACAGCAGTGATAGTGACAGCAGTGCACAGCAACGAAAGCAACGAATAGCAGTGAC	2805
Qy	1344	AGTAGTAAAAGC---GAATCCAATPAACTTTAACGGTAATTCGGTCAGTAGTAACATTTAC	1400
Db	2806	AGCAGTGATAGCAGCAACAGCAGTGTATAGTGACAGCAGTGATAGCAGCAACAGCAGTGAC	2865
Qy	1401	CGGTGATAGTACGGTAAACAGGTAGTATAGGTGGTACAGAACTTTCGCAACAGTGAAATAT	1460
Db	2866	ACCAGTGATAGCAGCAACAGCAGTGATAGCAGTGAACAGCAATATAGTAGTGACAACAGC	2925
Qy	1461	AGGAGCAGGAATAACATTACAGCGC	1486
Db	2926	AATAGCAGTGACAGCAACAGCAG	2951

```

RESULT 4
US-09-949-016-16237
; Sequence 16237, Application US/09949016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16237
; LENGTH: 12191
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16237

Query Match 3.1%; Score 59.6; DB 7; Length 12191;
Best Local Similarity 44.0%; Pred. No. 0.00014;
Matches 407; Conservative 0; Mismatches 504; Indels 15; Gaps 3;

```

```

: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN BT 474
: FILE REFERENCE: PB 0004 WO 8
: CURRENT APPLICATION NUMBER: US/10/203,138A
: CURRENT FILING DATE: 2002-08-02
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 04 February 2000 (04.02.00)
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 26 May 2000 (26.05.00)
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 03 August 2000 (03.08.00)
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 03 October 2000 (03.10.00)
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 27 September 2000 (27.09.00)
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 21 September 2000 (21.09.00)
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 30 June 2000 (30.06.00)
: NUMBER OF SEQ ID NOS: 15438
: SOFTWARE: Molecular Dynamics Sequence Listing Engine
: SEQ ID NO 7146
: LENGTH: 510
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL035457.11
: FEATURE:
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.6
: FEATURE:
: OTHER INFORMATION: NT HIT: U01287.1, EVALUATE 1.30e+00
: US-10-203-138A-7146

```

```

; Sequence 31796, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CU001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 31796
; LENGTH: 961
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-31796

Query Match      3.0%; Score 56.8; DB 9; Length 961;
Best Local Similarity 42.8%; Pred. No. 0.00043;
Matches 289; Conservative 0; Mismatches 387; Indels 0; Gaps 0.0043;

Qy 831 AGTAACCTGGAATATAGGTAATACAAATTCATTAGCCACCAATAGTGTAGGAGCAGGTAC 890
    || || || || || || || || || || || || || || || || || || || || ||
Db 227 AGCAACAACAACGACCAACAACAACAGCAGCAGCAGCAACAACAGTAGCACCACCACC 286
    || || || || || || || || || || || || || || || || || || || || ||
Qy 891 GGCTACGTTAGGGGGAGCGGTTATTAAGGCCACCTACGACTAAGTTAAGCAATGCGCGTC 950
    || || || || || || || || || || || || || || || || || || || || ||
Db 287 AACAGCAGCACACACACAGCAGCAGCAACAACAGCAGCAGCAACACAGCAGTAGCAACAAC 346
    || || || || || || || || || || || || || || || || || || || || ||
Qy 951 GGTATTAACTTACAAATGCAGTATTAAACAGGTGCTGTTGTAATACCAACAGCGCGTGA 1010
    || || || || || || || || || || || || || || || || || || || || ||
Db 347 AGCAGCAGCAACAACAACAGCAGTAAACAACAACAGCAGCAACAACAACAGCAATAGC 406
    || || || || || || || || || || || || || || || || || || || || ||
Qy 1011 TAATGTAGGTGTCGTAATCTTAAGTGGAGCATTGAGTCAAGTACACCGGGAATATAGGTAA 1070
    || || || || || || || || || || || || || || || || || || || || ||
Db 407 AGCAGCAACAACAGCAGCAGCAACAACAGCAGCAATTAACAACAACAGCAGCAATAACAC 466
    || || || || || || || || || || || || || || || || || || || || ||
Qy 1071 TACAAATTCATTAGCCAGCATTAATATAGGAGCAGGTGTGGCTACCTTTGGATGGACGGT 1130
    || || || || || || || || || || || || || || || || || || || || ||
Db 467 AATAACACAGCAGCAACAACAACAGCAGCAGCAACAACAGCAGCAACAGCAGCAGCAGC 526
    || || || || || || || || || || || || || || || || || || || || ||
Qy 1131 TATTAAGACTTACTACGACTAAGTTTAACAGATGATGCGTCAGTATTATATTTACGAATCC 1190
    || || || || || || || || || || || || || || || || || || || || ||
Db 527 AACAGCAGCAGCAACAACAACAGTAGTAGCAGCAGCAACAACAGCAGCAGCAACAACAGAGC 586
    || || || || || || || || || || || || || || || || || || || || ||
Qy 1191 TGTAGTAGTACCAGGCAATAGATATATACCGGTAAATGCCAATAAAGGTGTGGTAACTTT 1250
    || || || || || || || || || || || || || || || || || || || || ||
Db 587 AACAGCAGCAGCAGCAACAACAGCAACAACAGCAGCAGCAGCAACAACAGCAACAACAGCAG 646
    || || || || || || || || || || || || || || || || || || || || ||
Qy 1251 TACCGGAGCAGTAGTACGGTAACCGGTAATATAGSTAAACACGCGCAGTATTAGCAGAGGTAA 1310
    || || || || || || || || || || || || || || || || || || || || ||
Db 647 AGCAGCAGCAACAACAGCAACAACAACAGCAGCAGCAGCAACAACAGCAGCAGCAGCAAC 706
    || || || || || || || || || || || || || || || || || || || || ||
Qy 1311 CGTAGGAGCAGGTTTGCTGCAATATACAGCGGGAGTAGTAAAGCGGAATGCAATAAACTTT 1370
    || || || || || || || || || || || || || || || || || || || || ||
Db 707 AAGAGCAACACAGCAGCATTCAGCAGCAACAACAGCAGCAGCAACAACAGTAGCAACAACAGCAGC 766
    || || || || || || || || || || || || || || || || || || || || ||
Qy 1371 AACGGTAATGCGTCACTAGTAGTAACATTTACCGGTGATAGTACGGTAAACAGGTAGTATAG 1430
    || || || || || || || || || || || || || || || || || || || || ||
Db 767 AGCAACAACAATAGCAGCAGCAACAACAGCAGCAGCAGCAACAACAGCAGCAACAACAGCAG 826
    || || || || || || || || || || || || || || || || || || || || ||
Qy 1431 TGGTACAGAACTCTTCGCAACAGTGAATATAGGAGCAGGAATTAACATTTAGCAGCGGAGG 1490
    || || || || || || || || || || || || || || || || || || || || ||
Db 827 AGCAGCATCAACAGAGCAACAACAGCAGCAGCAGCAGCAACAACAGCAGCAACAACAGCAAC 886
    || || || || || || || || || || || || || || || || || || || || ||
Qy 1491 AAGCCTTAGCTGCGAAT 1506
    || || || || || || || || || || || || || || || || || || || || ||
Db 887 AACAGCAACAACAAAT 902
    || || || || || || || || || || || || || || || || || || || || ||

```

RESULT 6
US-10-144-771-31796


```

QY 616 ACAATAGGTAATTAATCCCGCAGCTCAATAAGCAATTTGGAGCAAGTACGCTTTCTCTT 675
Db 6347 GCGATAGCAGTACAGCAGCTAGTATAGCAGTACAGCAACAGCAGTACAGTACGCGATA 6406
QY 676 GGAGGGGCGAGTTATTAAAGCCACTAGCTAGTTAAACAATGCAAGTCCGCGGTATTACA 735
Db 6407 GCAGTGACAGCAGCAACAGCAGTACAGCTAGTATAGCAGTACAGCAGTACAGTATAGT 6466
QY 736 CTTACAAATGCAATGCAATTAACAGTGTCTGTTGATAACACACAGCGCGGTGATGAT 795
Db 6467 ACAGCAGCAACAGCAGTATAGCAGCAGCAGCAAGTACAGTACAGCAGTACAGCA 6526
QY 796 GTAGTGTCTTAATTTAAACGGAGCGTTGAGTCAAGTAACTGGAATATAGGTAAATACA 855
Db 6527 GCAACAGCAGTATAGCAGCAACAGCAGTATAGCAGTATAGCAGTACAGCAGTACAGT 6586
QY 856 AATTCATTAGCAGCAATAGTGTAGCAGCAGTACGCGTACGTTAGGGGGGCGGTTATT 915
Db 6587 GCGACAGCAGCAATAGCAGTACAGCAGTATAGTGTAGCAGCAGCGATAGCAGCAACA 6646
QY 916 AAAGCCACTACGACTTAAGTTAAACGAATGACAGCGTGGTATTAAACACTTACAAATGCA 975
Db 6647 GCAGTGATAGCAGCAGCAGCAGTACAGTACAGCAGTACAGCAGTACAGCAGCAATA 6706
QY 976 TTAACAGTGTCTGTGATAACACACAGCGCGGTGATATAGTGTGCGTAAATCTAAGT 1035
Db 6707 GAAGTGACAGTATAGTATAGTACAGCAGCAGTATAGCAGTACAGCAGCAACAGCAGT 6766
QY 1036 GGAGCATTGAGTCAAGTAACCGGGAATATAGTATACAAATTCATTAGCCACGATTAAT 1095
Db 6767 ACAGCAGTATAGTATAGTACAGCAGTACAGCAGCAAGCAGCAATAGCAGTACAGCA 6826
QY 1096 ATAGAGCAGTGTGGCTTACCTGGATGGAGCGGTTATTAAAGCTACTACGACTAAAGTT 1155
Db 6827 GTGATAGCAGCAACAGCAGTATAGTACAGCAGTATAGCAGCAACAGCAGTACAGCA 6886
QY 1156 ACAGATGATGCGTACGATTAATTTACGAATCCTGTAGTATGTTTAAACCGGAGCA 1209
Db 6887 GTGATAGCAGCAACAGCAGTATAGCAGTAAACGAGTATAGTATAGTACAGCAACAG 6946
QY 1210 ATAGATATACCGGTAATGCAATTAAGGTGTGTTGTTTACCGGAGCAAGTACGTTA 1269
Db 6947 GCAGTGACAGCAGCAACAGCAGTACAGCAGTATAGCAGTACAGCAGTATAGTATAGT 7006
QY 1270 ACCGATATATAGTATACAGCGCAGTATAGCAGCAGTATAGCAGTACAGCAGTGTG 1329
Db 7007 ACAGCAGCAATAGCGGTGACAGCAGCAACAGCAGTATAGCAGTATAGCAGTATAGC 7066
QY 1330 CAAATACAGCGGAGTAGTAAAGCGAATGCAATAAATTAACGGGATAATGCGTACGTA 1389
Db 7067 ACAGCAGTACAGCAGCAACAGCAGCAGTACAGCAGTACAGCAGTATAGCAGTACAG 7126
QY 1390 GT---ACATTTACCGGTATAGTACCGGTAAACAGGTATAGTGTGATAGCAGCACTTTC 1446
Db 7127 GTGACAGCAGTATAGCAGCAACAGCAGTATAGCAGTATAGCAGTATAGCAGTATAGC 7186
QY 1447 GCAACAGTGAATATAGCAGCAGCAATACATACAGCCG 1486
Db 7187 GCAGTATAGTATAGCAGCAGCAACAGCAGTACAGCAG 7226

```

RESULT 15

US-10-363-798-1/c

: Sequence 1, Application US/10363798

: GENERAL INFORMATION:

: APPLICANT: Kong, Xiangyin

: APPLICANT: Xiao, Shangxi

: APPLICANT: Zhao, Guoping

: APPLICANT: Yu, Chuan

: APPLICANT: Hu, Landian

: TITLE OF INVENTION: METHOD OF DIAGNOSING AND TREATING DENTINOGENESIS IMPERFECTA

: TYPE II USING DENTIN SIALOPHOSPHOPROTEIN GENE AND CODED

```

: TITLE OF INVENTION: PRODUCT THEREOF
: FILE REFERENCE: 9548.78USWO
: CURRENT APPLICATION NUMBER: US/10/363,798
: CURRENT FILING DATE: 2003-03-05
: PRIOR APPLICATION NUMBER: CN 00125042.6
: PRIOR FILING DATE: 2000-09-05
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1
: LENGTH: 8201
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-10-363-798-1

```

Query Match 2.88; Score 53.2; DB 8; Length 8201;

Best Local Similarity 41.1%; Pred. No. 0.0053;

Matches 379; Conservative 0; Mismatches 543; Indels 0; Gaps 0;

```

QY 102 TGGAGTCTCTTGGTCTGCAAGAACCGTAACCTGCTGATGGTGCAGAGCTTGCAGCGGAAC 161
Db 7459 TGCTGTCACTGCTGCTGCTGCTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7400
QY 162 AAATATAGTCTCGAGCGCGTCTTTTGTAGCGGGTTCTACTTTACAATATACCGGTGC 221
Db 7399 CACTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7340
QY 222 TTTTACGGTTACTGATGCTGACGTAAGTGTTCGTCGATTAAGTTAAATATTTTGCAGC 281
Db 7339 TGTCGCTGCTGCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7280
QY 282 AGGTCCTTTTTCAGTAATCTGCTGATATTTTCATTAGTTTCAGTGGTAGATACGGGAGGAGC 341
Db 7279 TGCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7220
QY 342 TAATAACTTTGCAGTTAATATTGATGATGTTTAAACCTTAACTTTACAGGTACCGGTAC 401
Db 7219 CACTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7160
QY 402 TGCAGCCTCAGGTCGCAATCCTGCTGTTGTTATTCCAAAGTGGACAAAGCTGCTGCTATAA 461
Db 7159 TATCACTGCTGTTGCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7100
QY 462 TACATATAGTCTTTTAGGTAATTAATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 521
Db 7099 TGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7040
QY 522 TTCAGATCCAGATGTTATTAGGACCAATAACGCTTTCAGGAGAAATATAGATGGAGGAGTAT 581
Db 7039 TGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6980
QY 582 AATACTGCAATACAGATGCTGCCATTAAACGGAACAATAAGTAACTAATCCGGCAGC 641
Db 6979 CACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6920
QY 642 TCAATAAGCATTTGGAGCAAGTACGCTTCTCTTTGAGGGGCGAGTTATTAAAGCCACTAC 701
Db 6919 TTTCACTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6860
QY 702 GACTAAGTTAAACAAATGTCAGCGCGGTATTAAACACTTACAAATGCAAAATGSCAGTATT 761
Db 6859 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6800
QY 762 AGGTGCTGTTGATAACACACAGCGGCTGATGATGATGATGATGATGATGATGATGATGAT 821
Db 6799 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6740
QY 822 GTTGAGTCAAGTAACTGGAATATAGTAAATTAACAAATTCATTAGCCGCAATAGTGTAG 881
Db 6739 CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6680
QY 882 AGCAGTACGGCTACCTTAGGGGAGCGGTTATTAAAGCCACTTACGACTTAAGTTAAAGAA 941
Db 6679 CACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6620

```

Qy 942 TCCACGTCGGTATTAACTTACAAATCCAGTATTAAACAGTCTGTGATACACCAC 1001
Db 6619 TATTACTGCTGCTACTGTATGTGCTGCTGCTATCAGTCTGCTGCTATCAGTCTGCTATCAGTGC 6560
Qy 1002 AGGCGGTGATAATGTAGGTGC 1023
Db 6559 TATCACTGCTGCTGCTATC 6538

Search completed: April 17, 2003, 19:41:14
Job time : 2298 secs

REFERENCE 3 (bases 1 to 805)

Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000)
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source

Location/Qualifiers

1..805
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="005C20"
 /clone_lib="H"
 /notes="Genoscope sequence ID : COBH005BB10XE1-end : PUC-ori"

BASE COUNT 87 a 194 c 293 g 226 t 5 others
 ORIGIN

Query Match 3.5%; Score 69.6; DB 17; Length 805;
 Best Local Similarity 50.6%; Pred. No. 5.9e-08;
 Matches 168; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 94 ACCGGTAGTGAGTCCTGGTGGCTGCAAGAACCGTAACCTGATGCTGCAGAGCTTGCA 153
 |||||
 Db 402 ACTGGTCTGCAGCTGGTGTGCTGCTACCTGGTGGTGGTGGTGGTGGTGGTGGTGG 461
 |||||
 QY 154 GCCGGAACAATAATPAGTCTCTGGACCGGCTGCTTTTGTAGCGGCTTCTACTTTACAATAT 213
 |||||
 Db 462 GCTGGTCTGCTACTGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 521
 |||||
 QY 214 ACCGGTCTTTTACGGTTACTGATGCTGACGTAAGTGTGCTGATGATGATGATGATGAT 273
 |||||
 Db 522 GCTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 581
 |||||
 QY 274 TTTCACAGCTCTTTTTCAGTAACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 333
 |||||
 Db 582 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 641
 |||||
 QY 334 GGAGGAGCTAATAAAGTTGTCAGTTAATATTGATGATGATGATGATGATGATGATGATGAT 393
 |||||
 Db 642 GCTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 701
 |||||
 QY 394 ACCGGTACTGCAGCTAGCTGCTGCAAACTCTCC 425
 |||||
 Db 702 GCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 733

RESULT 2

FR0008252
 LOCUS FR0008252 443 bp DNA linear GSS 02-MAR-1997
 DEFINITION F.rubripes GSS sequence, clone 190022bb9, genomic survey sequence.
 ACCESSION Z92062
 VERSION 292062.1 GI:1869276
 KEYWORDS GSS: genome survey sequence.
 SOURCE Takifugu rubripes.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 Tetraodontidae; Takifugu.

1 (bases 1 to 443)

Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrانيا, Y.,
 Williams, G. and Brenner, S.

Direct Submission

Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource

Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgm.mrc.ac.uk

Vector: pBluescript II KS

V_type: phagemid

PRIMER: KS

DESCR:

One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES

source

Location/Qualifiers

1..443
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone="190022bb9"
 /clone_lib="cosmid 190022"

BASE COUNT 38 a 111 c 144 g 143 t 7 others
 ORIGIN

Query Match 3.3%; Score 63.6; DB 17; Length 443;
 Best Local Similarity 47.8%; Pred. No. 2e-06;
 Matches 177; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 131 CTGCTGATGGTCAGAGCTTTCACCGGAAACAATAATAGTCTCTGGACCGGCTCTTTTG 190
 |||||
 Db 1 CGGCTGCTGAAGCNGTAGCTGCAGCTGCAGCTGTAGCTGTAGCTGTAGCTGTAGCTGTAG 60
 |||||
 QY 191 TAGCGGGTCTTACTTTACAATATACCGGTGCTTTTACGGTTACTGATGCTGCACGTAAGTG 250
 |||||
 Db 61 CTGCTGNTAGCTGTAGCTGCTGCTGTAGCTGTAGCTGTAGCTGTAGCTGTAGCTGTAG 120
 |||||
 QY 251 TTCTGTCATTAGATTTAAATAATTTTCAGCAGAGCTCTTTTTCAGTAACCTGGTGATATTT 310
 |||||
 Db 121 CTGCTGATGCTGAGCTGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
 |||||
 QY 311 CATPAGTTTCAGTGGTAGATACGGGAGAGAGCTAATAAAGTTTCAGTGTAAATATGATGATG 370
 |||||
 Db 181 CTGAGTGTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
 |||||
 QY 371 GTTTAACTTAACTTTAACAGGTACCGGTACTGCAGCTACCGGTGCAAACTCTGGTGT 430
 |||||
 Db 241 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 |||||
 QY 431 TATTCAGGTGGACAAAGCTGCTCTAATAATACATATAGTCTTTAGGTAATAATAACTC 490
 |||||
 Db 301 TAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
 |||||
 QY 491 TAGGTGGAGC 500
 |||||
 Db 361 TAGCTGCTGC 370

RESULT 3

BM588321/c
 LOCUS BM588321 711 bp mRNA linear EST 25-FEB-2002
 DEFINITION 17000687322518 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
 19600449697628 5', mRNA sequence.

ACCESSION BM588321

VERSION BM588321.1 GI:18884182

KEYWORDS EST.

SOURCE African malaria mosquito.

ORGANISM

Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.

1 (bases 1 to 711)

Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab
 R., Collins, F.H., Venter, J.C. and Hoffman, S.L.

Celera Anopheles gambiae EST project

Unpublished (2002)

Contact: Holt R.A.

Celera Genomics

45 W. Gude Dr.,

Rockville, MD 20850, USA

Tel: 2404533151

Fax: 2404534580

Email: HoltRA@celera.com

Plate: NU01004AAR. row: F column: 14

Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers

1..711
 /organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to Permethrin - std.
 chromosome)"

Db	33	GTTCATTTCTCGACATTT 14	
RESULT 5			
CNS0039C/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
1. .1101			
/organism="Drosophila melanogaster"			
/db_xref="taxon:7227"			
/clone="BACR08K10"			
/clone_lib="RPCI-98"			
/note="end : TET3"			
BASE COUNT	201 a	64 c	131 g
ORIGIN			
Query Match	3.2%	Score 61;	DB 17; Length 1101;
Best Local Similarity	16.9%;	Pred. No. 1.7e-05;	
Matches 119;	Conservative 290;	Mismatches 293;	Indels 3; Gaps 1;
QY	1123	GGAGCGTATTAACTACTAGCTAGTAAACAGATGATCGCTAGTATTAATTT 1182	
Db	1101	GKARRGGDTTWDHTRKDDWMTKWTWKKRADDERRWAGDARWMDWGAGTWTATW 1042	
QY	1183	ACGAATCCCTGTAGTACCGGACATAGATAACCGTAAATGCAATAAAGGTGTG 1242	
Db	1041	WWWATWDTWMDKWWATRAKTDTANTWRTANRADWAGDRGACKRRDRDAATDAG 982	
QY	1243	GTAATCTTACCGGACGAAGTACGGTAACCGGATAATAGTAAACCGGAGTATTACGA 1302	
Db	981	AGRRDGRKRKDKKDKGDDKGGKKKAAKAWATKWWDDWMDKDKWGDGAKDRK 922	
QY	1303	GAGGTAAAGCTAGGACAGTGTTCGCAATACACGCGGAGTAGTAAAGCGCAATGCA 1362	
Db	921	ADDDGAGDKDDGKGDADDTDGTDKDDDKDKWDDWMDKAGTWGATWAAATDWW 862	
QY	1363	ATAAATTAACGGATAATCGCTAGTAACTATTCACCGGTGATAGTACGGTAACAGGT 1422	
Db	861	GWADADWTTWDAADWDADWDADWAWKWDADWAWGARTADRDWGDGRGKRGAKRR 802	
QY	1423	AGTATAGGTGGTACAGACTCTTCGCAACAGTGAATATAGGAGCAGGAATTAACAT 1482	

/note="Vector: pME18S-FL3; Site_1: DraIII (CACCATGTC); Site_2: DraIII (CACTGTGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGCGCCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTC). xhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed and donated by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CATTCTGCTCTAAAGCTGCG and 3' end primer CGACCTGAGCTCGAGCACA."

BASE COUNT 218 a 234 c 37 g 152 t
ORIGIN

Query Match 3.1%; Score 60.4; DB 13; Length 641;
Best Local Similarity 44.7%; Pred. No. 1.9e-05;
Matches 235; Conservative 0; Mismatches 291; Indels 0; Gaps 0;

QY 97 GGTAGTGGAGTCTCTGCTGCAAGACCGTAACCTGCTGAGTTCAGAGCTTCAGCC 156
DB 582 GCTGGAGAGTCTCTGCTGCAAGACCGTAACCTGCTGAGTTCAGAGCTTCAGCC 523

QY 157 GGAACAATATAGTCTCTGAGCCGCTGCTTTTGTAGCGGTTCTACTTTACAATATAC 216
DB 522 GGAGAAGTGTAGCTGGAGAAGTTGTTGTGAAGTTGATGTAGCTGGAGAAGTTGTT 463

QY 217 GGTGCTTTTACGGTTACTGATGCTGACGTAAGTGTTCGTCGATGATTAATAATTT 276
DB 462 GTTGTGTTGAAGTTGATGTAGCTGGAGAAGTTGTTGTGTTGGAGAAGTTGATGTAGCT 403

QY 277 GCAGCAGGTCTTTTTCAGTAACTGCTGATATTTTCATTAGTTCAGTGTAGATCGGA 336
DB 402 GGAGAAGTGTGTTGTTGGAGAAGTTGATGTAGCTGGAGAAGTTGTTGTGTTGGAGA 343

QY 337 GGAGCTAATAAATCTCAGTTAATATTGATGATGGTTTAAACCTTAACCTTTAAGCAGTAC 396
DB 342 GTTGTATGATGTAGCTGGAGAAGTTGATGTTGTTGGAGAAGTTGATGTAGCTGGAGAAGTTGTT 283

QY 397 GGTACTGCAGCTACGGTCAATCTCGTGTGTTTATTCAGGTGGACAAGCTGCTGCT 456
DB 282 GTTGTGTTGAAGTTGATGTAGCTGGAGAAGTTGTTGTGTTGTTGAAGTTGATGTAGCTGGA 223

QY 457 AATAATACATATACTGCTTAGTATAATACTCTAGTGGAGCGAATGCCGTTTGACT 516
DB 222 GAAGTTGTTGTTGGAGAAGTTGTTGTTGGAGAAGTTGATGTAGCTGGAGAAGTT 163

QY 517 ATGCTTCAGATCCAGATGATTAAGACCAATAACGCTTGCAGGAATAATAGATGGAGGA 576
DB 162 GTTGTGTTGGAGAAGTTGATGTAGCTGAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCA 103

QY 577 GGTATTAATCAATCAATACAGATGCTGCCATTAACGGACAATAG 622
DB 102 GTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGTAG 57

RESULT 7
CNS033D4
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 208D22 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION
AL225985
VERSION
AL225985.1
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE
1 (bases 1 to 935)

AUTHORS
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 935)
AUTHORS
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL
Unpublished
REFERENCE
3 (bases 1 to 935)
AUTHORS
Genoscope.
TITLE
Submitted (12-APR-2000)
JOURNAL
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.

FEATURES
Location/Qualifiers
1..935
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="208D22"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG208DB11SP1-end : PUC-Ori"

BASE COUNT 136 a 251 c 273 g 271 t 4 others
ORIGIN

Query Match 3.1%; Score 60.4; DB 17; Length 935;
Best Local Similarity 47.2%; Pred. No. 2.3e-05;
Matches 184; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

QY 111 TGGTGCTGCAAGAACCGTAACCTGCTGATGTCAGAGCTTCAGCGGACAAATATAGG 170
DB 264 TGTAGTGTGCTGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 323

QY 171 TCGTGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 230
DB 324 AGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 383

QY 231 TACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 290
DB 384 AGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 443

QY 291 TTCAGTAACTGGTGATATTTTCATTAGGTTTCAGTGTAGATACGGGAGGAGCTAATAACT 350
DB 444 TCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 503

QY 351 TGCAGTTAATATTGATGATGGTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTT 410
DB 504 TCGTGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 563

QY 411 CGGTGCAAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 470
DB 564 TCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 623

QY 471 TCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 500
DB 624 TGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 653

RESULT 8
FR0036552
LOCUS
DEFINITION
Fugu rubripes GSS sequence, clone 032K14aF12, genomic survey sequence.
ACCESSION
AL124061
VERSION
AL124061.1
KEYWORDS
GSS; genome survey sequence.

```

SOURCE      Takifugu rubripes.
ORGANISM    Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1 (bases 1 to 536)
AUTHORS    Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrani,Y., Williams,G. and Brenner,S.
TITLE      Direct Submission
JOURNAL    Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hmp.mrc.ac.uk
COMMENT    Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESC:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES    Location/Qualifiers
            source
            1..536
            /organism="Takifugu rubripes"
            /db_xref="taxon:31033"
            /clone="032K14aF12"
            /clone_lib="cosmid 032K14"
BASE COUNT 44 a 136 c 165 g 157 t 34 others
ORIGIN
Query Match 3.1%; Score 60; DB 17; Length 536;
Best Local Similarity 47.3%; Pred. No. 2.2e-05;
Matches 192; Conservative 0; Mismatches 209; Indels 5; Gaps 1;
QY 110 TTGCTGCTGCAAGAACGGTAAGTCTGCTGATGCTGAGAGCTTGCAGCGGACAAATATAG 169
Db 41 TAGCTGCTGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAG 100
QY 170 GTCCTGAGCGGGTCTTTGTAGCGGGTCTACTTTACAAATATACACGGTGTCTTTACGG 229
Db 101 CTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 160
QY 230 TTACTGATGCTGAGCTGAAGTGTTCGTCATTAGATTTAAATAATTTTGCAGCAGGCTTTT 289
Db 161 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAG 220
QY 290 TTTGAGTAACTGATGATGATTTTCAATAGTTCAGTGTAGATACGGAGGAGCTAATAAAC 349
Db 221 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAG 280
QY 350 TTGAGTAAATATGATGATGATTTTCACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 404
Db 281 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 340
QY 405 AGCTACGGTGCAGTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 464
Db 341 TGCTGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 400
QY 465 ATATACGCTTTAGGTAATATTAATCTAGGTGGAGCGAATGCCGGT 510
Db 401 TGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 446

RESULT 9
AF046361/c AF046361 Mus musculus 129Sv/Ev Mus musculus genomic clone OSTI5546,
LOCUS      454 bp DNA linear GSS 02-NOV-2000
DEFINITION DNA sequence.
ACCESSION AF046361
VERSION    AF046361.1 GI:3005232
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE   1 (bases 1 to 454)
AUTHORS    Zambrowicz,B.P., Friedrich,G.A., Buxton,E.C., Lilleberg,S.L.,
            Person,C. and Sands,A.T.
TITLE      Disruption and sequence identification of 2,000 genes in mouse
            embryonic stem cells
JOURNAL    Nature 392 (6676), 608-611 (1998)
MEDLINE    98219085
COMMENT    Contact: Zambrowicz BP
            OmniBank
            Lexicon Genetics Incorporated
            4000 Research Forest Drive, The Woodlands, TX 77381, USA
            Email: materials@lexgen.com
            OmniBank Sequence Tag
            Class: exon-trapped.
FEATURES    Location/Qualifiers
            source
            1..454
            /organism="Mus musculus"
            /strain="129Sv/Ev"
            /db_xref="taxon:10090"
            /clone="OSTI5546"
            /clone_lib="Mus musculus 129Sv/Ev"
            /cell_type="embryonic stem cell"
BASE COUNT 155 a 104 c 94 g 70 t 31 others
ORIGIN
Query Match 3.1%; Score 59.6; DB 17; Length 454;
Best Local Similarity 51.7%; Pred. No. 2.6e-05;
Matches 138; Conservative 0; Mismatches 123; Indels 6; Gaps 1;
QY 189 TGTAGCGGTTCTACTTACAAATATACCGGTGCTGCTTTACGGTTACTGCTGACGTAAAG 248
Db 295 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 236
QY 249 TGTTCGTGCATTAGATTTAAATAATTTTGCAGCAGGCTTTTTTTCAGTAACTGCTGATAT 308
Db 235 TGNFCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 176
QY 309 TTCAATTAGTTCAGTGTAGATACGGAGGAGCTAATAAACTTGCAGTTAATATTATGATGA 368
Db 175 TTNTGTTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122
QY 369 TGGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 428
Db 121 TGTCTTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 62
QY 429 GTTATTTCAAGGTGGACAAGCTGCTGC 455
Db 61 NNCAGNCAAGATAAAGNACCGCTCC 35

RESULT 10
AW128683/c AW128683 819 bp mRNA linear EST 07-JUN-2001
LOCUS      fa36a05.y1 zebrafish WashU MPMG EST Danio rerio cDNA clone
DEFINITION IMAGE:3740912.5' similar to TR:Q920W6 Q920W6 PAX TRANSCRIPTION
            ACTIVATION DOMAIN INTERACTING PROTEIN PTIP. ;contains element TARI
            repetitive element ;, mRNA sequence.
ACCESSION  AW128683
VERSION     AW128683.1 GI:6116587
KEYWORDS   EST.
SOURCE     zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostarlophysi; Cypriniformes
            ; Cyprinidae; Danio.
            1 (bases 1 to 819)
REFERENCE   1 (bases 1 to 819)
AUTHORS    Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
            WashU Zebrafish EST Project 1998
            TITLE

```

JOURNAL
COMMENT

Unpublished (1998)
Other ESTs: fe36a05.x1
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 485.

FEATURES
source

BM630395/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

BM630395
17000687501370 A.Gam.ad.cdna1 Anopheles gambiae cDNA clone
19600449647565 5', mRNA sequence.
BM630395
BM630395.1 GI:18929906
EST.
African malaria mosquito.
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab
, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU010049WA row: A column: 03
Seq primer: M13 Reverse.
Location/Qualifiers
1..699
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449647565"
/clone_lib="A.Gam.ad.cdna1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."
BASE COUNT 261 a 230 c 155 g 53 t
ORIGIN

Query Match 3.1%; Score 59.2; DB 10; Length 819;
Best Local Similarity 47.5%; Pred. NO. 4.6e-05;
Matches 209; Conservative 0; Mismatches 228; Indels 3; Gaps 1;
QY 97 GGTAGTGGAGTCCTTGGTGCAGACCGTAAGTGTGATGGTGACAGCTTCACGCC 156
DB 589 GCTGCTGATTCCTGCTGTTGTTGACGCTGTGATGCTGCTGCTGCTGCTGCT 530
QY 157 GGAACAAATATAGGTCCTGGAGCGGTGCTTTTGGAGCGGTCTCTACTTACATATACC 216
DB 529 GTAGGAATGCTGTGCTGCTGCAAAATGCTGTGGTGCATCTGCTGCTGCTGCTGCT 470
QY 217 GGTGCTTTTACGGTTACTGCTGACGTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCT 276
DB 469 GCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 410
QY 277 GCAGCAGCTCTTTTTCAGTAACTGCTGATATTCATTAGTTCAGTGGTACAGCGGA 336
DB 409 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 350
QY 337 GGAGCTAAACCTTGCAGTTAAATATTCATGATGTTTAACTTAACTTAACTGCTGCT 396
DB 349 GTTGAACCTGCACTGAAGT---CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 293

QY 397 GGTACTGCAGCTACGGTCCAAATCCTCGCTTGTATTTCACAGGTGGACAAGCTGCTGCT 456
DB 292 CATGTTGCTGCTGATGCTGCAAGGCGAGCTCGGAAGCTGTTGGGGTCCATTTGTTGTG 233
QY 457 AATAATACATATACCTCTTTAGGTATAATACTCTAGGTGGAGCGAATGCCGGTTTGACT 516
DB 232 GGAATGCATGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 173
QY 517 ATTGCTTCAGATCCAGATGT 536
DB 172 ACATCTGCTGATGCATCTGT 153
RESULT 11
BM630395/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

BM630395
17000687501370 A.Gam.ad.cdna1 Anopheles gambiae cDNA clone
19600449647565 5', mRNA sequence.
BM630395
BM630395.1 GI:18929906
EST.
African malaria mosquito.
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab
, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU010049WA row: A column: 03
Seq primer: M13 Reverse.
Location/Qualifiers
1..699
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449647565"
/clone_lib="A.Gam.ad.cdna1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."
BASE COUNT 261 a 230 c 155 g 53 t
ORIGIN

Query Match 3.0%; Score 58.4; DB 13; Length 699;
Best Local Similarity 49.7%; Pred. NO. 7e-05;
Matches 149; Conservative 0; Mismatches 151; Indels 0; Gaps 0;
QY 125 CCGTAACCTGCTGATGCTGCAGAGCTTCGACGGGCAACAAATAGTCTCGAGCGCGTG 184
DB 421 CTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 362
QY 185 CTTTGTAGCGGGTCTCTACTTTACAATATACCGGTCTTTTACGGTTACTGATGCTGAGC 244
DB 361 CTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302
QY 245 TAAGTGTTCGTGCAATTAGATTTAAATAATTTTGCACGAGGCTCTTTTTCAGTAACTGGT 304
DB 301 TAGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 242

Qy	342	TATAAAGCTTCGACCTTAATTATGTGGTG	370
Dd	122	TGCTGCTAGTCTGTTGCTGCTGCTGCTG	94
<hr/>			
RESULT	13	BG837033	857 bp mRNA linear EST 25-MAY-2001
BG837033/c	Zm08_10H09_A	Zm08_AAFC_ECORC_Fusarium_graminearum_inoculated_corn_ear_Zea_mays	
LOCUS	cDNA clone Zm08_10h09	mRNA sequence.	
DEFINITION	GB837033	GB837033.1 GI:14203356	
ACCESION	VERSION	KEYWORDS	EST.
SOURCE	ORGANISM	Zea mays. Zea mays. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 857) Harris,L.J., Balcerzak,M., Allard,S., Saparno,A., Couroux,P., De Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S., Singh,J.A, Sprott .D. and Tinker,N.A. Expressed Sequence Tags from Developing Maize Kernels Six Days after Silk Channel Inoculation with Fusarium graminearum Unpublished (2001) Contact: Harris, Linda J. Eastern Cereal and Oilseed Research Centre Agricuture and Agri-food Canada Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6, CANADA Tel: (613) 759-1314 Fax: (613) 759-6566 Email: harrislj@em.agr.ca. Location/Qualifiers 1. 857 SOURCE	
REFERENCE	AUTHORS		
TITLE	JOURNAL		
COMMENT			
FEATURES			

```

source
I. .85/
/organism="Zea mays"
/cultivar="CO430"
/db_xref="taxon:4577"
/clone="Zm08_10h09"
/clone_lib="Zm08_AAFC_ECORC_Fusarium_graminearum_inoculate
d_corn_ear"
/tissue_type="Developing kernels (sibcrossed)"
/dev_stage="10-11 days post-silk emergence"
/notes="Vector: Bluescript SK+/XhoI-ECORI; Site_1: ECORI;
Site_2: XhoI; Field-grown maize ears were silk
channel-inoculated in the morning (~10 am) with 1 ml of a
Fusarium graminearum macroconidial suspension (500,000
spores/ml) and whole ears were collected and immediately
frozen in liquid nitrogen 6 days later."
BASE COUNT      273 a      220 c      217 g      146 t      1 others
ORIGIN

Query Match      3.0%; Score 57.8; DB 12; Length 857;
Best Local Similarity 48.3%; Pred. No. 0.00011;
Matches 161; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

Qy 169 GGTCTGTGAGCGCGTGTCTTTTGATCGCGGGTCTACTTTACAATATACCGGTGCTTTTACG 228
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 GTTCTGCTGCTGCTGCATTGCTGCATCTGTTCTGTTGAGGCTGCATTGTGCTGCTGCT 384
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 229 GTTACTGATGCTGACGTAAGTGTTCGTGCATTAGATTTAAATAATTTTGCAGCAGGTCTT 288
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 383 GCTGCTGCATTGCTGTGCATCTGTTGCTGCTGAGGCTGCATTGCTGTGCTGCTGCTGCA 324
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 289 TTTTCAGTAACTGGTGATATATTCATTAGGTTTCAGTGGTAGATACGGAGGAGCTTAATAAA 348
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 TTTGCTGCACTGTGTGCTGCTGAGGCTGCATCTGTTGCTGCTGCTGCTGCTGCTATTT 264
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 349 CTTGCAGTTAATATGATGATGGTTTAAACCTTAACCTTTTAACAGGTACCGGTACTGCAACC 408
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 GCTGCATCTGTGTGTTGATGCTGCATCTGTTGCTGTGTGATGCTGATCTGCTGCTGCA 204
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Qy	409	TACGGTGCAAATCTCGGTGTTATTCCAGGTGGACAAGCTGCTGCTAATAFACATAT	468
Dd	203	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	144
Qy	469	ACTGCTTAGTAATAAAGCTCTAGGTGGAGCG	501
Dd	143	GCTGCATCTGGAAGGTGCTGCTGTCATCG	111
RESULT 14			
CNS05HXN			
LOCUS			
DEFINITION			
ACCSSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches 175; Conservative			
Qy	181	GGTGCTTTTGACGGGTTCTACITACATAATACCGGTGCTTTTACCGTTACTCATGCT	240
Dd	636	GCTGCTACTGTTGCTGCTACTGTTGCTGCTACTGTTGCTGCTGCTGCTGCTGCTGCT	755
Qy	241	GACGTAAGTGTTCGTCATTAGATTAATAATTTGCAGCAGGTCCTTTTTCAGTAACT	300
Dd	756	GCTGCTACTGTTGCTGCTACTGTTGCTGCTACTGTTGCTGCTGCTGCTGCTGCTGCT	815
Qy	301	GGTGATATTTCAATAGGTTTCAGTGGTAGATACGGGAGGAGCAATFAAACCTTTCAGTAACT	360
Dd	816	ACTGTTGCT-----GCTACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	869

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:00:54 ; Search time 40 seconds
(without alignments)
2062.053 Million cell updates/sec

Title: US-09-800-065-2

Perfect score: 3022

Sequence: 1 MANISLKLFQRAIQGLKTA.....MIELSLKIIYPVLLTVVVS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	A_Geneseq_101002:*
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3022	100.0	619	AAE10890	Rickettsia felis o
2	333.5	11.0	1026	AAE10890	Caulobacter S-layer
3	331	11.0	1026	AAW37490	Caulobacter cresce
4	331	11.0	1026	AAAY44757	Caulobacter cresce
5	323.5	10.7	1026	AAAR48993	rsaa S-lyaeer prote
6	322	10.7	1180	AAAB01845	Haemophilus influe
7	322	10.7	1188	AAAB01844	Haemophilus influe
8	316.5	10.5	1222	AAAB01830	H. influenzae stra
9	316.5	10.5	1228	AAAB01828	Haemophilus influe
10	310	10.3	1095	AAAB01835	Haemophilus influe

11	310	10.3	1101	21	AAAB01834	Haemophilus influe
12	303	10.0	1267	11	AAAR04232	Rickettsia rickett
13	299.5	9.9	1338	14	AAAR41731	High molecular wei
14	293.5	9.9	1598	18	AAW30291	Non-typeable Haemo
15	298.5	9.9	2732	22	ABB52855	Escherichia coli p
16	298.5	9.9	1529	14	AAAR41732	High molecular wei
17	298.5	9.9	1601	18	AAW30292	Non-typeable Haemo
18	293.5	9.7	969	21	AAAB01827	Haemophilus influe
19	293.5	9.7	975	21	AAAB01826	Haemophilus influe
20	292	9.7	1612	19	AAW65088	R. prowazekii S-la
21	290.5	9.6	1221	21	AAAB01825	Haemophilus influe
22	290.5	9.6	1227	21	AAAB01824	Haemophilus influe
23	286	9.5	1477	14	AAAR41724	High molecular wei
24	280	9.3	1036	21	AAAB01849	Haemophilus influe
25	280	9.3	1477	14	AAAR41728	High molecular wei
26	280	9.3	1477	21	AAAB01848	Haemophilus influe
27	279	9.2	1477	18	AAW30294	Non-typeable Haemo
28	278	9.2	1477	15	AAAR63506	Haemophilus high m
29	276	9.1	871	21	AAAY95550	Chlamydia pneumoni
30	276	9.1	1407	23	ABB90541	Chlamydia pneumoni
31	273.5	9.1	1095	21	AAAB01847	Haemophilus influe
32	273.5	9.1	1536	14	AAAR41723	High molecular wei
33	273.5	9.1	1536	14	AAAR41725	High molecular wei
34	273.5	9.1	1536	15	AAAR63505	Haemophilus high m
35	273.5	9.1	1536	21	AAAB01846	Haemophilus influe
36	273	9.0	1749	22	AAAB72314	Glutamic acid rich
37	272.5	9.0	1536	18	AAW30293	Non-typeable Haemo
38	269	8.9	1005	21	AAAB01833	Haemophilus influe
39	269	8.9	1011	21	AAAB01832	Haemophilus influe
40	263	8.7	1648	23	ABB54925	Lactococcus lactis
41	262	8.7	998	21	AAAB01843	Haemophilus influe
42	262	8.7	998	21	AAAB01842	Haemophilus influe
43	260	8.6	943	19	AAW81745	M. tuberculosis im
44	260	8.6	943	19	AAW64378	Mycobacterium tube
45	260	8.6	943	20	AAAY39175	M. tuberculosis an

ALIGNMENTS

RESULT 1
AAE10890
ID AAE10890 standard; Protein; 619 AA.
XX
AC AAE10890;
XX
DF 18-DEC-2001 (first entry)
XX
DE Rickettsia felis outer membrane protein (rompA).
XX
KW Outer membrane protein; rompA; antibacterial.
XX
OS Rickettsia felis.
XX
PN WO200166691-A2.
XX
PD 13-SEP-2001.
XX
PF 06-MAR-2001; 2001WO-US07820.
XX
PR 06-MAR-2000; 2000US-187323P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Bouyer DH, Crocquet-Valdes P, Stenos J, Walker DH;
XX
DR WPI; 2001-596829/67.
XX
DR N-PSDB; AAD18232.
XX
PT Novel isolated Rickettsia felis outer membrane protein and
XX
PT polynucleotide for modulating expression of the protein in a host cell

appicant

PS Claim 27; Page 62-64; 73pp; English.

CC The present invention relates to an isolated Rickettsia felis outer membrane protein (rompA) and its polynucleotide. The rompA gene is useful for modulating expression of the protein in a host cell. The rompA gene is useful as probe or for the design of primers to obtain DNA encoding the protein by either cloning and colony/plaque hybridisation or amplification using PCR. The present sequence is Rickettsia felis outer membrane protein (rompA).

XX
SQ Sequence 619 AA;

Query Match 100.0%; Score 3022; DB 22; Length 619;
Best Local Similarity 100.0%; Pred. No. 2.9e-186;
Matches 619; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MANISKLKFORAIQKGLKLTFTSTAAILMTGSGVLTADGAELAGTNGIPGA 60
Db 1 MANISKLKFORAIQKGLKLTFTSTAAILMTGSGVLTADGAELAGTNGIPGA 60

Qy 61 GAFVAGSTLQVTGFTVDADVSRALDNNFAAGLFSVTGDISLGSVVDGFGANKLAVN 120
Db 61 GAFVAGSTLQVTGFTVDADVSRALDNNFAAGLFSVTGDISLGSVVDGFGANKLAVN 120

Qy 121 IDDLGLTLTGTGAAYGANPALLFOGGAAANNNTYALGNITLGGANAGLTIASDPDVL 180
Db 121 IDDLGLTLTGTGAAYGANPALLFOGGAAANNNTYALGNITLGGANAGLTIASDPDVL 180

Qy 181 GPITLAGNIDGGIITDNTDAINGTNTGNTNPAQAISIGASTLSLGGAVIKATTTKLTNA 240
Db 181 GPITLAGNIDGGIITDNTDAINGTNTGNTNPAQAISIGASTLSLGGAVIKATTTKLTNA 240

Qy 241 APVLTLTNANAVLTGAVDNTTGGDDVGLNUNGALSQVTGNIGNTSLATISVGAGTATL 300
Db 241 APVLTLTNANAVLTGAVDNTTGGDDVGLNUNGALSQVTGNIGNTSLATISVGAGTATL 300

Qy 301 GGAVIKATTTKLTNAASVLTITNAVLTCGAVDNTTGGDNGVNVNLSGALSQVTGNIGNTNS 360
Db 301 GGAVIKATTTKLTNAASVLTITNAVLTCGAVDNTTGGDNGVNVNLSGALSQVTGNIGNTNS 360

Qy 361 LATINIGAGVATLDCGAVIKATTTKLTDDASVLIFTPVVTGAIIDNTGNANKGVVIFTGA 420
Db 361 LATINIGAGVATLDCGAVIKATTTKLTDDASVLIFTPVVTGAIIDNTGNANKGVVIFTGA 420

Qy 421 STVTDNIGNTAVLAESVSGAGLLQIQGVVKANINLTDNASVYFTFGDSTVTSIGGTE 480
Db 421 STVTDNIGNTAVLAESVSGAGLLQIQGVVKANINLTDNASVYFTFGDSTVTSIGGTE 480

Qy 481 LFAVTNIGAGITLRAGGSLAANNIDFGAASNLFEENGAGKNYNLIGTIANGNNATLINA 540
Db 481 LFAVTNIGAGITLRAGGSLAANNIDFGAASNLFEENGAGKNYNLIGTIANGNNATLINA 540

Qy 541 AGTVIANDVSGTVAQINQNNKIFVINAKNADVDILDAQAISFKGAASRLFLANVSLOM 600
Db 541 AGTVIANDVSGTVAQINQNNKIFVINAKNADVDILDAQAISFKGAASRLFLANVSLOM 600

Qy 601 IELSLKLIIPVLLTVVVS 619
Db 601 IELSLKLIIPVLLTVVVS 619

RESULT 2
AAR94014
ID AAR94014 standard; Protein; 1026 AA.
XX
AC AAR94014;
XX
DT 21-MAY-1996 (first entry)
XX
DE Caulobacter S-layer protein.
XX
KW S-layer; rsaA gene; surface protein; biofilm; vaccine.
XX

OS Caulobacter crescentus strain CB15.
XX
FH Key Location/Qualifiers
FT Region 860..905
ET /note= "glycine-aspartic acid repeat region"
XX
PN US5500353-A.
XX
PD 19-MAR-1996.
XX
PF 09-JUN-1992; 92US-0895367.
XX
PR 09-FEB-1994; 94US-0194290.
XX
PR 09-JUN-1992; 92US-0895367.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI Bingle WH, Smit J;
XX
DR WPI; 1996-171046/17.
XX
DR N-PSDB; AAT17717.
XX
PT Expression and presentation of polypeptide heterologous to
PT Caulobacter S-layer protein to bacterium's environment - by cloning
PT in-frame into Caulobacter rsaA gene, and expressing as fusion prod.
XX
XX with S-layer protein
PS Example 2; Fig 6a-c; 22pp; English.
XX
CC The Caulobacter crescentus CB15 paracrystalline S-layer protein
CC (AAR17717) is encoded by the rsaA gene (AAR17717). It can be used as
CC a fusion partner with polypeptides of interest, allowing
CC presentation of the polypeptide on the surface of the Caulobacter
CC cells. The bacterium is cultured as a biofilm in a bioreactor
CC or may be used to present an antigenic epitope (see AAR94016 and
CC AAR94018) to the environment e.g. for use as a vaccine.
XX
SQ Sequence 1026 AA;

Query Match 11.0%; Score 333.5; DB 17; Length 1026;
Best Local Similarity 24.7%; Pred. No. 2.9e-13;
Matches 181; Conservative 76; Mismatches 232; Indels 243; Gaps 33;

Qy 16 GLKTALETTTAAIMLTGSGVL-----GAARTVTAD-GAELAAGT-----NIGP 58
Db 315 GIETMNVTSAAITLNTSSGVTGLTALNTNTSGAAQVTAGAGONLTATTAOAAANNVAV 374
Qy 59 GAGAFV-----AGSTLOYTGFTVDADVSRALDNNFAAGLFSVTGDISL----- 105
Db 375 DGGAVTVASTGVTSGTFT--VGANSAACTSVSVANSSTTTTGAIVTGGTAVTVTAQT 432
Qy 106 -GSYVD-----TGGANKLAVNIDDLGLTLTGTGTAAYGANPALLFOGGQAAANN 154
Db 433 AGNAVNTLTQADVTVTGNSTTAVTVQ--TAAATAGATVAGRVNGAVTITDSAAASAT 490
Qy 155 TYTALGNITLGGANAGLTIASDPDVLGPITLAG-----NIDG 191
Db 491 TAGKIATVTLGSFGAA-TI--DSSALTTVNLSGTSLGIGRGALTATPANTLTNLVNG 547
Qy 192 ---GGIITDNTDAA-----INGTIGNTPAAQISIGASTLSLGGAVIKATTTKLTNA 240
Db 548 LTTTGAITDSEAAADGGFTTINAGSTIASLVAADATTNLISGDARVTTTSHTAAA 607
Qy 241 APVLTLTNA-----NAVLTGAVDNT----- 260
Db 608 LTGITVNSVGTALGAELATGLVFTGAGRDSILLGATTKAIVNGAGDDTVTVSSATLGA 667
Qy 261 ----TGSD--DVGVLNUNGALSQVTGNIGNTSLATISVGAGTATLG-----G 302
Db 668 GGSVNGGSDGTDVLVANVNGSSFPAGFGFETLRV---AGAAQGSNHNANGFTALQLG 723
Qy 303 AVIKATTTKLTNA-----SVLTLTNAVLTGAVDNTTGGDNGVNVNLSGALSQVTGNIGN 357
Db 303 AVIKATTTKLTNA-----SVLTLTNAVLTGAVDNTTGGDNGVNVNLSGALSQVTGNIGN 357

Db 724 ATAGATT--FTNVAVNGLTVLAAPTGTTTVTILANATGSDVFNLTSSSAAALAGTVA- 780
 QY 358 TNSLATINICA---GVATLDGAVIKATYTKLTDDASVLIFNPNVVTGCAIDNTGNANKG 413
 Db 781 LAGVETVNIATDNTNTAHVDLTTLQATSAK-----SIVVTG-----NAG 820
 QY 414 VVIFTGASTVTIGNIGTAVLAESVAGLLQIQGGVVKANALNLTDNASVVTFTGDSVTI 473
 Db 821 LNL-----INTGNTA-----VTSFASAVTGTGSATVFSANTIV 855
 QY 474 GSIGGTELPATVNIAGITLRAGGSLAANNIDFGA-ASNLEFNGPAGKYNLIGTIANGN 532
 Db 856 GEV-----VTIRGGAGADSLTGSATANDTIIGGAGADTLVVTGGT-----DFTTGGT 902
 QY 533 NATL-NINAGT-----VIANDVSIGTVAQINQNNKIFVINAKNADVDILDAQAISFKGA 587
 Db 903 GADIFDINAIGSTAFVTITDAVG-----DKLDLVLGISTNGA 940
 QY 588 -ASRLFLANVSL 598
 Db 941 IADGAFGAATVTL 952

RESULT 3

AAW37490
 ID AAW37490 standard; Protein; 1026 AA.

AC AAW37490;

XX 20-APR-1998 (first entry)

XX Caulobacter crescentus S-layer rsaa protein.

XX S-layer; rsaa gene; Caulobacter; vaccine; antigenic; ligand; enzyme;
 KW metallothionein; heavy metal; water; sewage; xylanase; cellulase;
 KW wood pulping.

XX Caulobacter crescentus.

XX WO9734000-A1.

XX 18-SEP-1997.

XX 10-MAR-1997; 97WO-CA00167.

XX 12-MAR-1996; 96US-0614377.

XX (UYBR-) UNIV BRITISH COLUMBIA.

XX Bingle WH, Nomellini JF, Smit J;

XX WPI: 1997-470880/43.

XX N-PSDB; AAV01866.

XX New DNA containing sequence for C-terminal region of Caulobacter
 PT S-layer protein - expressed as fusion proteins containing antigenic
 PT peptides in Caulobacter, useful as live vaccines

XX Example 2; Fig 6; 58pp; English.

XX The present sequence represents a Caulobacter S-layer protein used in an
 CC example of the present invention. A new DNA construct has been developed
 CC which contains at least one restriction site for insertion of DNA
 CC upstream of DNA encoding a C-terminal region of at least the last 82
 CC amino acids (aa) of Caulobacter S-layer protein. Caulobacter containing
 CC the DNA constructs above additionally containing a sequence encoding a
 CC heterologous polypeptide, are particularly useful in live vaccines
 CC (where the heterologous polypeptide is an antigen). They can also be
 CC used for production of e.g. ligands, enzymes or other proteins, e.g.
 CC metallothioneins to remove heavy metals from water or sewage, or
 CC xylanase or cellulase for use in wood-pulping. All known Caulobacter
 CC strains are harmless, and stable in outdoor environments, including
 CC water (so suitable for vaccinating fish) or soil. They are well suited

CC for growing in biofilm reactors and produce S-layer proteins, which is
 CC an ideal system for presentation of antigens, at high level.

XX Sequence 1026 AA;

XX Query Match 11.0%; Score 331; DB 18; Length 1026;

XX Best Local Similarity 27.3%; Pred. No. 4.2e-13;

XX Matches 181; Conservative 70; Mismatches 202; Indels 210; Gaps 36;

QY 23 TTSTRAIMLTGSGVLAART-----VTADG-----AELAACTNIGPGA 60

Db 413 TTTTGAIAVTCGTAVTVAOTAGNAVNTTLTQADVITVGTSSSTAVTIVTQATAAT-----A 467

QY 61 GAFVAGSTLOYTCAFTVTDADVS-----VRALDLNNEAAGL-----FSVTG-- 101

Db 468 GATVAG---RVNGAVTITDASAASATTAGKIATVTLGSFGAATIDSSALTVNLSTGTS 524

QY 102 -DISLGSVVDTTGGANKLAVNIDDLTLTGTGTAYGANPALLFOGGOAANNNTYALG 160

Db 525 LGIGRGALTATPTANTLTNLVNG---LTTTGALT-----DSEAAADDGFT-- 566

QY 161 NITLGGANAGLIAS-----DPDVLGPITLAGNIDGG--IITDNTDAAING-TIGTNP 213

Db 567 TINIAGSTASSTIASLVAADATTL-----NISGDARVITTSHTAAALTGITVTVNSVGA 619

QY 214 ---AQISIGASTLSLGA-----VIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGD- 264

Db 620 TLGAELATG--LVFTGGAGRDSILLGATTKAIVMGAGDDTVTVSSATL-GAGSVNNGDG 676

QY 265 -DVGVNLNGALSQVGTGNIGTNSLATISVGAGATPLG-----GAVIKATTTK 311

Db 677 TDVLVANVNGSFSADPAFGFETLRV---AGAAAGGSHNANGFTALQLGATAGATT-- 730

QY 312 LFNAA-----SVLTTLNAVLTGAVDNTTGGDNVGVNLSGALSQVGTGNIGTNSLATINI 366

Db 731 FTNVAVNGLTVLAAPTGTTTVTILANATGSDVFNLTSSSAAALAGTVA-LAGVETVNI 789

QY 367 GA---GVATLDGAVIKATTTKLTDDASVLIFTNPPVVTGAIDNTGNANKGVVIFTGAST 422

Db 790 AATDNTTAHVDTLTLQATSAK-----SIVVTG-----NAGLNL----- 823

QY 423 VTDNIGNTAVLAESVSVGAGLLQIQGGVVKANALNLTDNASVVTFTGDSVTGTSIGTETF 482

Db 824 --TNTGNTA-----VTSFASAVTGTAPAVTFVSANTTVGEV----- 858

QY 483 ATVNIGAGITLRAGGSLAANNIDFGA-ASNLEFNGPAGKYNLIGTIANGNATL-NINA 540

Db 859 VTIRGGAGADSLTGSATANDTIIGGAGADTLVVTGGT-----DFTTGGTGAIDFIDNA 911

QY 541 AGT---VIANDVSIGTVAQINQNNKIFVINAKNADVDILDAQAISFKGA-ASRLFLAN 595

Db 912 IGTSFAVVTITDAVG-----DKLDLVLGISTNGAIDAGAFGAA 949

QY 596 VSL 598

Db 950 VTL 952

RESULT 4

AAW44757

ID AAY44757 standard; Protein; 1026 AA.

XX AAY44757;

XX 04-MAY-2000 (first entry)

XX Caulobacter crescentus surface layer protein.

XX Surface layer protein; S-layer secretion signal; antibiotic; vaccine;
 KW recombinant fusion protein cleavage; enzyme; protein polymer;
 KW antibacterial enzyme; foodstuff.

XX Caulobacter crescentus.

XX Key Location/Qualifiers
 FT Cleavage-site 692..693
 FT /note- "Asp-Pro dipeptide present in S-layer secretion
 FT signal sequence. It is a site where a fusion
 FT protein comprising a target protein and the secretion
 FT signal is cleaved"
 XX MO200004170-AL.
 PN XX
 PD 27-JAN-2000.
 XX 14-JUL-1999; 99WO-CA00637.
 PF 14-JUL-1998; 98CA-2337704.
 PR (UYBR-) UNIV BRITISH COLUMBIA.
 XX Smit J;
 XX WPI; 2000-182434/16.
 DR N-PSDB; AAZ50079.
 XX
 XX Cleavage of Caulobacter produced recombinant fusion proteins useful for
 PT producing vaccine peptides
 PT
 XX
 PS Example 1; Pages 21-23; 33pp; English.
 CC The patent discloses a method for cleaving a recombinant fusion protein
 CC which is produced by Caulobacter and consists of Caulobacter surface
 CC layer (S-layer) protein (containing the C-terminal secretion signal) and
 CC a target protein heterologous to Caulobacter. The cleavage of target
 CC protein from the S-layer protein is carried out under mild acid
 CC conditions so that cleavage occurs at aspartate-
 CC proline dipeptide site without solubilising the protein.
 CC The cleavage is accomplished while the fusion protein is in
 CC an insoluble aggregate form which facilitates purification of the
 CC protein. The method is useful for producing pure proteins including
 CC recombinant human and animal therapeutic antibiotic and vaccine peptides,
 CC enzymes, protein polymers, and antibacterial enzymes for foodstuffs.
 CC The present sequence is a S-layer protein from C. crescentus.
 CC The S-layer secretion signal, corresponding to the C-terminal portion of
 CC the protein from amino acid 690 onwards, is fused with a target sequence
 CC for construction of a recombinant fusion construct which is expressed in
 CC Caulobacter.
 XX
 XX Sequence 1026 AA;
 SQ
 Query Match 11.0%; Score 331; DB 21; Length 1026;
 Best Local Similarity 27.3%; Pred. No. 4.2e-13;
 Matches 181; Conservative 70; Mismatches 202; Indels 210; Gaps 36;
 QY 23 TTSTAAMLTGSGVLGAART-----VTADG-----AELAAAGTNIGPGA 60
 DB 413 TTTTGAITAVTGGTAVTAGNAVNLTLLQADVTVTGNSSTAVTQTAAAT-----A 467
 QY 61 GAFVAGSTLQYTGFTVDVDS-----VRALDLNFAAGL-----FSVTG-- 101
 DB 468 GATVAG---RVNGAVTTDSAAASATPAGKIATVTLGSGAATTIDSSALTTVNLSTGCT 524
 QY 102 -DTSLSGWSWDTGGANKLAVNIDGLTLTGTGTAAGANPALLFOGQAAANNVTYALG 160
 DB 525 LGIGRGALTATPANTLTNLVNG---LTTTGAIT-----DSEAAADGFT--- 566
 QY 161 NITLGGANAGLTAS-----DPDVLGPITLAGNIDGG--IITDNTDAING-TIGTNPA 213
 DB 567 TINIAGSTASSTIASLVAADATTL-----NISGDARVITTSHTAAALGIIVTSVGA 619
 QY 214 ---AQISIGASTLSLGA-----VIKATTKLNAAAPVLITNANVLTAAGVNTTGGD- 264
 DB 620 TLGAELATG--LVFTGGAGRDSILLGATTAIVMGAGDDTIVTVSSATL-GAGGSVNGGDG 676
 QY 265 -DVGVLNGLALSQVTGNIGNTNSLATISVGAGTATLGT-----GAVIKATTTK 311

DB 677 TDVLVANVNGSFSADPAFGGFETLRV-----AGAAAGSHNANGFTALQLGATAGATT-- 730
 QY 312 LTNA-----SVLTLTNAVLGTADVNTTGGDNVGNLNGALSQVNTGNIGNTNSLATINI 366
 DB 731 FTNVAVNVGLTVLAAPTGTITVTLANATGSDVFNLTSSSAAALAGTVA-LAGVETVNI 789
 QY 367 GA-----GVATLDGAVIKATTTKLTDDASVLFTNPVVVVTGAIDNTGNANKGVIFTGAST 422
 DB 790 AATDTNTTAHVDTLTTLQATSATK-----SIVVTG-----NAGLNL----- 823
 QY 423 VTDNIGNTAVLAESVCGAGLLQIOGGVVVKANAINLTNANASVVTGSDTSTVTSCTGTELF 482
 DB 824 --TNTGNTA-----VTSFDSAVTGTCTAPAVTEVSANTTVGEV----- 858
 QY 483 ATVNIGAGITLRAGSLSAANNIDFGA-ASNLFEENGPAKNYNLIGTTIANGNNATL-NINA 540
 DB 859 VTRGGAGADSLTGATANDTIIGGAGADTLVYTGCT-----DTFTGGTGADIFDINA 911
 QY 541 AGT-----VIANDVSIGTVAQINIONNKFVINAKNADVDIILDAQAISFKGA-ASRLFLAN 595
 DB 912 IGTSTAFVTITDAVVG-----DKLDLVGISTINGAIDAGAFGA 949
 QY 596 VSL 598
 DB 950 VTL 952
 RESULT 5
 AAR48993
 ID AAR48993 standard; Protein; 1026 AA.
 XX AAR48993;
 AC AAR48993;
 DT 14-SEP-1994 (first entry)
 XX rsaa S-lyaeer protein.
 DE
 XX C. crescentus; rsaa; paracrystalline; S-layer; protein; heterologous;
 KW cellulase; xylase; metallothionein; restriction site;
 KW reading frame; fusion protein; bioreactor; toxic metal; sewage;
 KW waste water; wood pulp suspension; cell surface; vaccine; fish.
 XX Caulobacter crescentus.
 OS
 XX CA2090549-A.
 PN
 XX 10-DEC-1993.
 PD
 XX 26-FEB-1993; 93CA-2090549.
 PF
 XX 09-JUN-1992; 92US-0895367.
 PR
 XX (UYBR-) UNIV BRITISH COLUMBIA.
 PA
 XX Bingle WH, Smit J;
 PI
 XX WPI; 1994-066249/09.
 DR
 XX N-PSDB; AAQ57972.
 DR
 XX Prodn. of heterologous polypeptides in bacteria, partic.
 PT Caulobacter - by expression of a fusion prod. of the polypeptide
 PT sequence and a bacterial S-layer protein gene
 PT
 XX Claim 17; Fig 6; 27pp; English.
 PS
 XX This sequence is encoded by the C. crescentus rsaa gene and represents
 CC the paracrystalline S-layer protein. The rsaa gene was used in the
 CC production of the heterologous protein of the invention. The
 CC heterologous protein is produced by cloning a polypeptide coding
 CC sequence, eg. cellulase, xylase or a metallothionein, into a
 CC restriction site within the rsaa gene which preserves the rsaa reading
 CC frame and expressing the fusion sequence in Caulobacter. This S-

CC layer protein bacterial system can be used in bioreactors, eg. to
 CC bind toxic metals in sewage waste water etc. or for the treatment of
 CC wood pulp suspensions. The system can be used to produce heterologous
 CC proteins at the cell surface for use in vaccines, partic. fish
 CC vaccines. The S-layer protein is synthesised in large quantities and
 CC has a general repetitive sequence, permitting the synthesis of large
 CC amounts of heterologous protein as a fusion product and presentation at
 CC the cell surface.
 XX
 SQ Sequence 1026 AA;

Query Match 10.7%; Score 323.5; DB 15; Length 1026;
 Best Local Similarity 24.8%; Pred. No. 1.3e-12;
 Matches 181; Conservative 78; Mismatches 232; Indels 239; Gaps 34;

QY 16 GLKTLFTSTALMLTGSGLV-----CAARTVTAD-CAELAACT-----NIGP 58
 DB 315 GIETMNTSGAALTNTSSGVTGLTALNTNTSGAATVTAGAGQNLTTATTAQAANNVAV 374
 QY 59 GAGAFV-----AGSTLQYTGAFVTDDADVSVRALDNLNFAAGLFSVTGDLSL-----NIGP 105
 DB 375 DGGANVTASTGVTSGTIT--VCANSAAGSTVSVSANSTTTTGAIVTGGTAVTVAQT 432
 QY 106 -GSVD-----TGANKLAVNIDGLTLTLTGTCTAAYGANPALLFOGQAAANN 154
 DB 433 AGNAVNTTLTQADVTVTGNSTTAVTQ--TAAATAGATVAGRVNGAVTITDSAAASAT 490
 QY 155 TYTALGNITLGGANGLTTASDPDLVPITLAG-----NIDG 191
 DB 491 TAGKATVTLGSGFAA-TI--DSSALTVNLSGTGTSLGIGRALTATPTANTLTNLVNG 547
 QY 192 ---GGIITDNTAA-----INGTIGNTPAAQISIGASTLSLGGAVIKATTKTLTNA 240
 DB 548 LTTGATITDSEAADGFTTINAGTASSTIASLVAADATTILNISGDARVITTSHTAA 607
 QY 241 APVLITNA-----NAVLTGAVDNT-----260
 DB 608 LTGITVNSVGTALGAEATGLVFTGGAGADSILLGATTKAIVMGAGDDTVTVSSATLGA 667
 QY 261 ----TGDD--DVGVLNLGALSQVTGNIGNTNSLATISVAGATATLG-----G 302
 DB 668 GGSVNGDGTDLVAVNNGSSFSADPAFGFPETLRV----AGAAAQGSNANGFTALQLUG 723
 QY 303 AVIKATT-TK--LTNAASVLTITNAVLTAVDNTTGGDNVGVNLSGALSQVNTGNITGN 359
 DB 724 ATAGATTFKVAVNVGLTVLAAPTGTITVLANATGTSDFENLTLSAALAAGTVA-LA 782
 QY 360 SLATINIGA---GVATLGAIVKATTTKLTDDASVLITFNPVVVGTGADTGNANKGVV 415
 DB 783 GVETVNIATDNTTAAHVDTLTLQATSAR-----SIVVTG-----NAGLN 822
 QY 416 IFTGASTVTDNIGNTAVLAESVVGAGLLQIOGGVVKANALNTDNASVVTFTGDSVTGCS 475
 DB 823 L-----TNGNTAVTS-----FDASAVTGTGRA--VTFVSANTVTGE 857
 QY 476 IGGTELFATVNICAGITLRAGSLAANNIDFGA-ASNLEFNGPAGKNYNLIGTIANGNA 534
 DB 858 V-----VIRGAGADSLTGSATANDTIIIGAGADTLVYTGST-----DFTTGTGA 904
 QY 535 TL-NINAAGT---VIANDVSGTVAQINQNNKIFVINAKNADVDILDAQISFKGA-A 588
 DB 905 DIFDINAIGTSTAFVITITDAVVG-----DKLDLVGISTNGAIA 942
 QY 589 SRLFLANVSL 598
 DB 943 DGAFGNAVTL 952

RESULT 6
 AAB01845
 ID AAB01845 standard; Protein; 1180 AA.
 XX
 AC AAB01845;

XX 11-SEP-2000 (first entry)
 DE Haemophilus influenzae strain 15 mature HMW2A protein, SEQ ID NO:65.
 XX
 KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
 KW non-typeable Haemophilus influenzae; NTHI; non-encapsulated;
 KW recombinant production; Escherichia coli; antibacterial; vaccine;
 KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
 KW detection; diagnosis.
 XX
 OS Haemophilus influenzae strain 15.
 XX WO2000020609-A2.
 PN 13-APR-2000.
 PD 07-OCT-1999; 99WO-CA00938.
 PF 07-OCT-1998; 98US-0167568.
 PR 08-DEC-1998; 98US-0206942.
 PA (CONN) CONNAUGHT LAB LTD.
 XX Loosmore SM, Yang Y, Klein MH;
 DR WPI: 2000-303789/26.
 DR N-PSDB; AAA52194.
 XX
 PT Nucleic acid molecule for producing recombinant high molecular weight
 PT proteins of Haemophilus which are used as a vaccine to provide
 XX protection against Haemophilus induced diseases in humans -
 PS Claim 8; Flg 27A-R; 307pp; English.

CC The invention relates to the recombinant production of Haemophilus
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
 CC expression construct used to effect recombinant expression comprises a
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
 CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
 CC influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene
 CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
 CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
 CC and the hmwB and hmwC genes encode accessory proteins which are
 CC responsible for post-translational processing and secretion of the HMW
 CC proteins. The modified hmwABC operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMW. The invention also discloses hmwA genes (AAA52175-A52198)
 CC and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae
 CC strains J95c, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. influenzae HMW
 CC proteins which can be used as vaccines to mediate a humoral or
 CC cell-mediated immune response to provide protection against diseases in
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
 CC antigens in immunoassays for detecting antibodies against Haemophilus,
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
 CC HMW proteins can be used to isolate and clone hmw genes from other
 CC non-typeable strains of Haemophilus via hybridisation reactions. The
 CC present sequence represents a mature HMW protein from a non-typeable
 CC strain of H. influenzae.

SQ Sequence 1180 AA;

Query Match 10.7%; Score 322; DB 21; Length 1180;
 Best Local Similarity 25.2%; Pred. No. 1.9e-12;
 Matches 159; Conservative 83; Mismatches 251; Indels 138; Gaps 30;

QY 48 AELAACTNIGPGAGF-VAGSTLOYTGATV-----TDADVSVRALDNLNFAAGLFSVT 100
 DB 543 AEIQGNISQREGNLTSSDKINTKTIKKGVNGENSDDSTKS-----QANITKIT 596
 QY 101 GDISLGSVYDTGGANKLAVNIDDLTLTGTCTAAYGANPALLF-----GGQA 150


```

Db 990 GSIGGNAAV--TLNLTGALTYKSSINANS---GTLVINAKDAELNGEASGHTTV 1044
QY 526 GTIANGNNATLINAAGTVIANDVSTGTVAQINIQNNKIFIVINAKNADVDILDAQAISPK 585
Db 1045 -----NAT-NANGSGVIAT-----TSSRVNITGD-LITINGLN-----IISK 1080
QY 586 GAASRLFLANVSLQMIELSLLKIIVPLLVTV 616
Db 1081 NGINTVLLKGVKID-----VKYIQGIASV 1105

```

RESULT 8

AAB01830

ID AAB01830 standard; Protein; 1222 AA.

AC AAB01830;

DT 11-SEP-2000 (first entry)

DE H. influenzae strain K1 mature full-length HMW1A protein, SEQ ID NO:37.

KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;

KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;

KW recombinant production; Escherichia coli; antibacterial; vaccine;

KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;

KW detection; diagnosis.

OS Haemophilus influenzae strain K1.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match

Best Local Similarity 10.5%; Score 316.5; DB 21; Length 1222;

Matches 153; Conservative 92; Mismatches 257; Indels 141; Gaps 26;

QY 42 TVTADGAELAAAGTNIIGPGAGAF-VAGSTLQYTGAFVTADVSRLDLNN-FAAGLFSV 99

Db 536 TNGNDNTEIQIGNISQKEGNTLTISSDKVNITERIFI-KAGVNGDSDSNEATSANLTIK 594

QY 100 TGDISLGSVVDGCGANKLAVNIDDDGLTLTLTGTAAYCANPALLFQGGQAANNVYDAL 159

Db 595 TKELKLTNDLNTSGFNKAEITAKDNSNLT-----GDSNAGNTDAK- 636

QY 160 GNITLGGANAGLTIASDPVLGPITLAGNIDGGIITDAAINGTIGNTNPAAOISIG 219

Db 637 -KVTFSNVSKISASDHNV-----TLNSKVETSG-DTSDTEGNNNTGLITAKNVTVN 690

QY 220 ASTLSLGGGAVIKATTKLTNAAPVLTLTNANAVLTGAVDNTTGGDDVGVNLNGALSOVT 279

Db 691 NNITSHKVTNITASENVTTKAGTTINATGTSVEVTAKTGDIKGGIESGNVNITASGDT 750

QY 280 GNIGN-TNSLAISVAG-TATLGGGAVIKATTKLTNAASVLTNAV-----LTGAVD 331

Db 751 LNVSNITGQNVTVAAASGAVTTKGTINAT---TGNANITTKTGEINGEYKVSAGNVN 806

QY 332 NTTGGDNVGVNLGALSOVTGNTGNTNSLATINIGAGVATLDGAVIKATTKLTDDASV 391

Db 807 ITASGNTLVNSNITGQNVTVTANS-----AITTEGSTINAT---TGDANI 850

QY 392 LIFTNPVVVTGAIDNTGNANKGVV--IFTGASTVTDNIGNTAVLAESVVGAGLLQIQ--- 446

Db 851 -----TTQTGNGINGKVESSSGSVTLIATGQTLAVGNISGDTV--TITADKGLTTQTS 902

QY 447 -----GGVVKANALINDNASVVTFTGD-----STVTSIGG 478

Db 903 KINGTKSVTSSQSGDISGTSISGTVSVSATSLTQTQAGSKIEAKTGEANVTSATGIG 962

QY 479 TELFATVNI-----GAGITLRA-GGSLAANNIDFGAASNLEFNGPAG 519

Db 963 TISGTVNVTVANTDNLTIKDGARIKATGAVTLTATGGLTFTTSSDITSSNGQTLTA- 1021

QY 520 KYNLIGTIANGNNATLINAAGTVIAN---DVSIGTVAQINIQNNKIFIVINAKNADVD- 575

Db 1022 KDSSTAGSI-NAANVTLNTTGLTTVAGSKIEAASGTL-----VINAKAQLDG 1069

QY 576 -----ILDAQAISPKGA--ASRLFLANVSLQMIELSLLKII 609

Db 1070 AASGDHTVVVNATNANGSGSVIATTSRVNITGDLITINGLNII 1112

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 9

AAB01828

ID AAB01828 standard; Protein; 1228 AA.

AC AAB01828;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

recombinant production; Escherichia coli; antibacterial; vaccine; human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis; detection; diagnosis.

Haemophilus influenzae strain K1.

Key Location/Qualifiers
Misc-difference 313 /note= "Encoded by GG"

W0200020609-A2.
13-APR-2000.
07-OCT-1999; 99WO-CA00938.
07-OCT-1998; 98US-0167568.
08-DEC-1998; 98US-0206942.
(CONN-) CONNAUGHT LAB LTD.
Loosmore SM, Yang Y, Klein MH;
WPI: 2000-303789/26.
N-PSDB: AAA52179.

Nucleic acid molecule for producing recombinant high molecular weight proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans -
Claim 12; Fig 20A-R; 307pp; English.

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwaBC operon from a non-typeable (non-encapsulated) H. influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene clusters termed hmwiABC and hmwaBC. Each hmwaBC operon comprises hmwa, hmwb and hmwc genes. The hmwa genes encode the structural HMWA proteins and the hmwb and hmwc genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwaBC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwa genes (AAA52175-A52198) and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae strains Joyce, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone hmw genes from other non-typeable strains of Haemophilus via hybridisation reactions. The present sequence represents an HMWA protein from a non-typeable strain of H. influenzae.

Sequence 1228 AA;
Query Match 10.58; Score 316.5; DB 21; Length 1228;
Best Local Similarity 23.88; Pred. No. 4.4e-12;
Matches 153; Conservative 92; Mismatches 257; Indels 141; Gaps 26;

42 TVTADGAELAAAGTNGPGAGAF-VAGSTLQYTGATVTDADSVRALDLNN-FAAGLFSV 99
542 TNNGDNTEIQGGNISKEGNLTSSDKVNITERITI-KAGVNGDSDSNEATSANLTIK 600
100 TGDISLGSVDTGGANKLAVNIDPLTLTGTGTAAAGANPALLFGQGAANNNTYAL 159
601 TKELKLTNDLNLISGFNKAETAKDNSNLT-----GDNSDAGNTDAK- 642

160 GNITLGGANAGLTITASDPDLGPTTLAGNIDGGGIITDNDAAINGTIGNTPAAQISIG 219
643 -KVTFSSNVKDSKISASDHN-----TLNSKVETSG-DTSDTEGGNNNTGLTITAKNVTN 696
220 ASTLSLGGAVIKATTTKLTNAAPVLTLTNANAVLTGADVNTTGGDDVGVNLNGLALSQVT 279
697 NNITSHKVTNITASENVTTRAGTINATGSGVEVTAKTGDIKGGIESNGVNITASGDT 756
280 GNIGN-TNSLATISVGAG-TATLGGAVIKATTTKLTNAASVLTLTNAV-----LTGAVID 331
757 LNVSNITGQNVTVAAASGAVTTTKGSTINAT-----TGNANITTKTGEINGEVKSASGVN 812
332 NTTGGDNVGVNLGSGALSGQVTGNTGNINSLATINIGAGVATLGDGAVIKATTTKLTDDASV 391
813 ITASGNTLNVSNITGQNVTVTANS-----AITTEGSTINAT-----TGDANI 856
392 LIETNPVVVTGAIDNTGNANKGVV-IFTGASTVTDNIGNTAVLAESVSGAGLLQIQ--- 446
857 -----TTQGNINGKVESGSGVTLTATGQTLAVGNISGDTV--TTADKGLTKTTQSS 908
447 -----GGVVKANAILTNDNASVVTFTGD-----STVTGSIGG 478
909 KINGTKSVTTSSQSGDISGTISGNTVSVSATGSLTTQAGSKIEAKTGEANVTSATGTIGG 968
479 TELFATVNI-----GAGITLRA-GGSLAANNIDFGAASNLFEFGPAG 519
969 TISGNTVNVNTANTDLNLTGKARIKATGGAVTLTATGTLTETSSDITSSNGQTTLTA- 1027
520 KYNLIGTIANNNATNINAAGTVIAN--DVSIGTVQAQINQNNKIFVINAKNADVD- 575
1028 KDSSIAGSI-NAANVTLTGTLTVAGSKIEAASGTL-----VINAKDAQLDG 1075
576 -----ILDAQATISFKGA--ASRLFLANVSLQMIELSLKII 609
1076 AASGDHTVWATNANGSGSVIATTSRRVNTIGDLITINGLNII 1118

RESULT 10
AAB01835
ID AAB01835 standard; Protein; 1095 AA.
XX AAB01835;
XX
XX 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain LCDC2 mature HMWA protein, SEQ ID NO:45.
XX Mature HMW protein; hmw gene; hmwa1; hmwa2; high molecular weight;
XX non-typeable Haemophilus influenzae; NTHI; non-encapsulated;
XX recombinant production; Escherichia coli; antibacterial; vaccine;
XX human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX detection; diagnosis.
XX
XX Haemophilus influenzae strain LCDC2.
XX
PN W0200020609-A2.
XX
XX 13-APR-2000.
XX
XX 07-OCT-1999; 99WO-CA00938.
XX
XX 07-OCT-1998; 98US-0167568.
XX 08-DEC-1998; 98US-0206942.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Yang Y, Klein MH;
XX
XX WPI: 2000-303789/26.
XX DR N-PSDB; AAA52184.
XX
XX Nucleic acid molecule for producing recombinant high molecular weight
XX proteins of Haemophilus which are used as a vaccine to provide

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwABC operon from a non-typeable (non-encapsulated) H. influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwA genes (AA52175-A52198) and HMWA proteins (AAB01824-H01949) from the non-typeable H. influenzae strains Joyn, K1, K21, LDCDC2, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW

proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by *H. influenzae* (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against *Haemophilus*, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone *hmw* genes from other non-typeable strains of *Haemophilus* via hybridisation reactions. The present sequence represents an HMWA protein from a non-typeable strain of *H. influenzae*.

Sequence	1101	AA:
SQ		

Query Match	10.3%	Score 310;	DB 21;	Length 1101;
Best Local Similarity	23.1%	Pred. No. 1e-11;		
Matches 174; Conservative	96;	Mismatches 240;	Indels 242	

QY	18	KTALFTSTRAIMLTGSGV-----LGAARTVTADGAELAAAGTNIGPGAGA	62
Dd	356	RPTAFIESDLTLNATCGNISLNQVAGIDGNLQKLSIVANKNITFEGGNTILAAKKP----	411
QY	63	FVAGSTILOYTCAPFVTD-ADVSVRALDNLNFAAGFSVGTDI-SLGSVVVDTCGANKLAVN	120
Dd	412	--LEIKGNTIVKEGANVTLRSANYGNDKSAL-SIRGNVTNKGNLTVTGS--AIN	460
QY	121	IDGLTITLCTGTAAAYCANPALFFOGQAANNVYTALGNITLGGAN-----AGLTI	173
Dd	461	IERNLTV-----EGSAKFALNPYSFNVSUGLFDNCKNSI-SIAKGGAHFKDINTKSLNI	515
QY	174	ASPDVLGPITLAGNI---DGGGIITONTDAA----INGTI---GN-----TNPAAQI	216
Dd	516	TTSDSAYRTHIEGNTITSNGDLNITDKNNAEIQIGNISQKEGNTISSDKINITNQI	575
QY	217	SI-----GASTLSLGAVIKATTTKLT-----	238
Dd	576	TIKGVNKEDSDSTANNANLTTKTRELQTLDLINSIFDKAETAKEGADLIIGNSDN	635
QY	239	---NAAPV-----LTLTANNAVLTGAVDNTGGDDVGVLNIGA---LSQVTGN	281
Dd	636	NNAKKVTFENQVKDSKISADSHNVTLNSKVEYSNGNDAESNNGDGTSLTINAKNITVN	695
QY	282	IGNTNSLATISVA--GTATLGGAVIKATTTKLTJNAAASVLTJNNAVLTGAVDNTTG---	335
Dd	696	-NNTISHKVTNITASENVTTKAGTTINATGSVE---VTAKTGDIKGVESTSGSVTL	749
QY	336	---GDNVGVVNLSSALSOVTGNTGNTNSLATINIGACVATLDGAVIKATTTKLTDDRASVL	392
Dd	750	TATGEALANSISGNTVITIANKGK---LTOAGSTVPSAING---VTASSQSGDISGT	801
QY	393	IFTPNPVVVTGAIDNTGNANKGVVIPTCASVTODIGNTAVLAEVSVGAGLLQIOGGVVKA	452
Dd	802	ISCNTVKVSAIGDLTKSGSEIKAKTEANVTSATGTI-----GGTISG	845
QY	453	NAINLTDNASVVFDTGDSV-----	472
Dd	846	NAVNVATAN-----TGDLTWEDAADKIDATGGAATLTATSCLTKTKASSITSANNQVNL	899
QY	473	--TGSGTTELFA-TVNIGAGITLRAGSLEANNDIFGA---ASNLEPNGAGKNYNL	524
Dd	900	AKDGSIGGINANAVTLNTTGALTUVKGSINANS---GTLVINAKDAELNGEASGHNV	956
QY	525	IGTIANGNNATLINAAGTVIANDVSIGTVAQINOQNKFIVINAKNADVIDLDAQAISF	584
Dd	957	V-----NAT-NANGSGSVIAT-----TSSRVNITGD-LITINGLN-----IIS	992
QY	585	KGAASRLFLANVSLQMIELSLLKIIPVLLTV	616
Dd	993	KNGINTVLLKGVKID-----VKYIQOPGASV	1018

RESULT 12
AAR04232
ID AAR0

ID AAR04232 standard; protein; 1267 AA.

XX	AA04232;
AC	
XX	
XX	17-DEC-2001 (updated)
DT	14-SEP-1990 (first entry)
DT	
XX	Rickettsia rickettsii p130 and 27.6kD surface proteins.
DE	
XX	
XX	Rocky Mountain Spotted Fever; vaccine; diagnosis.
KW	
XX	
OS	Rickettsia rickettsii R strain.

Accession	Key	Location/Qualifiers
FM000001	Protein	/label=120 kD surface protein
FM000002	Protein	596...1267
FM000003	Protein	/label=27.631 kD surface protein

PN USN7429936-N.

13-MAR-1990

XX
PF :11-NOV-1989.

11 - NOV - 1980 - 80000 04200000

XX

XX

XX
XX

XX
MFT, 1930-1939/10/10.

PT Rickettsia rickettsii surface protein gene - used for diagnosis and for producing protein for vaccines against Rocky Mountain Spotted Fever.

PS Disclosure; ; Opp; English.

Main disclosure is the DNA fragment encoding the 120kD surface protein and the 726bp ORF encoding the 27.6 kD protein, however the specification best available copy is of too poor a quality to produce a beneficial index file.

Sequencing of the p120 gene showed an ORF starting with ATG at position 226. The ORF has 3900 bp which encode 1300 amino acids resulting in a protein with deduced mol.wt. of 132.611 kD. The amino acid sequence has 31 potential N-glycosylation sites. The other ORF of 726 bp was found about 100 bp downstream and on the opposite strand of the p120 gene. Between the two ORF's is the 106 bp intergenic region which presumably serves as a transcriptional termination site.

CC The recombinant protein can be used to vaccinate humans against RMSF. The
CC DNA can be used as a hybridisation probe for diagnosis of RMSF.

CC (Note: Revised entry submitted to correct the patent number format of
CC US Government-owned NTIS applications to prevent clashes with ongoing US
CC granted patent numbers. For further information please visit the Derwent
CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)

Sequence 1267 AA;

Query Match 10.0%; Score 303; DB 11; Length 1267;
Best Local Similarity 23.6%; Pred. No. 3.4e-11;

9 FQKAIQKGLK-TALFTTSTAAMLTGSGVLG-----AARTVTADGELAAGTN 55
 14 FAMIVQVGADGTTAFRTAASKVITITODSNFGVNTDFGNLAAQIKVANAITLTGNGTGDASN 73

QY 56 IGPAG--AFVAGSTLQYTGAFVTVDADVSVRALDLNFAAGLEFSVTGDISLGSVVDTGG 113

Db 74 PGNTAGVITEDANGTLESASA-----DANVAV-----TNNTITAT-----FASACACUOISC 120

09 114 ANKI AVNIIDRCI ETI ETI TCTCTCEAAVCANDRAY EACCCOAA ANATMUTET GNTMUT GCR 109

+ + + + +
THE FOLLOWING INFORMATION IS UNCLASSIFIED
DATE 05-11-2010 BY 60322 UCBAW

QY 168 N-----AGLTI-----ASDPDLGPITLAGNIDGGII-----TDNTDAINGT 206
 Db 169 NIICAGCTTOLCANGCTIKLTSTNNIVYDFLAIATDGTGVVDASSLTNAOTLTINGK 228
 QY 207 IGWTFNPAQISIGASTLSLGGAVIKATTKLTNAAPVLTITNANAVLTGAVDNTT-----261
 Db 229 IH-----TIGANKTKLGTQIGTSKTVLSNGVAINLVICNGDGAVFANDTTLITRT 280
 QY 262 -----GGDV--GVNLNGLSALSOVTGNIGN-TNSLA-----TISVAGT-----297
 Db 281 TNAAGGKIIFXPVN-NGTLAAGT-NGSATNPLAGINTSGVGVVDVTLNVEGVNLYA 338
 QY 298 -----ATLGGAVIKATTK-----LTNAASVLTITNNAVLTG-----328
 Db 339 TNITTTDANVGSFVFNASSNIYSGTVGGQGNKFNTVALENGCTTKFLGNATFNGNTTI 398
 QY 329 -----AVDNTGGD--NVGVNLS-----GALSOVT-----352
 Db 399 AANSTLOISGNYTADCVASADGTGVEFVNTGPITVTLNKEAAPVNALKOITVSGPNV 458
 QY 353 -----GNIGNTNSLAT-----INIGAVATLDG-----AVIKAT-----TTKLTLD 387
 Db 459 INEIGNAGNNGAVTDTIAFNSLGAUVFLLPGIPENDAGNTMPLIKSTVGNTAKGPD 518
 QY 388 -----DASVLTNPVVVTGAIDNTGNANKGVVIFTGASTVTDNIGNTAVL 433
 Db 519 VPSVVLGVDSVADGQVIVDQNNVIGLIGSDNGIIVNATTLYAGISTLNN-----571
 QY 434 AEVSAGLLOIOGGV--VKANAINLTDNASVVTFTGDSVTGTSIGGTLPAT-VNIGAG 490
 Db 572 -----OGTVTLGGVPNTPGTVYGLSTISASKFKGVFTTGTNNLSNIATNATINDG 625
 QY 491 ITRAGSLAANNIDFCAASNLBNFNGPAGKNYNLIGTIANGNNATNLINAAGTVIANDVS 550
 Db 626 VTVTGGI-----AGIGIGFDGI-----TLGSV-NGNG--NVPFADGILSNST 667
 QY 551 I-----GTVAQIQNNKIFVINAKNADVDILDQAISFKGAASRLFL-ANVSLQMI 601
 Db 668 MICTKANNGTVYL-----GNAFVGNIGSDTPV---ASVRFETGSGAGLOGNIVSQVI 720
 QY 602 E 602
 Db 721 D 721
 RESULT 13
 AAR41731
 ID AAR41731 standard; Protein; 1338 AA.
 XX
 AC AAR41731;
 XX
 DT 26-APR-1994 (first entry)
 XX
 DE High molecular weight protein 3 (HMW3).
 XX
 KW HMW; high molecular weight protein; virus; vaccine; influenza;
 KW epitope; immunity; haemophilus influenzae.
 XX
 OS Haemophilus influenzae.
 XX
 PN WO9319090-A.
 XX
 PD 30-SEP-1993.
 XX
 PF 16-MAR-1993; 93WO-US02166.
 XX
 PR 16-MAR-1992; 92GB-0005704.
 XX
 PA (BARE/) BARENKAMP S J.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Barenkamp SJ;
 XX

DR WPI; 1993-320683/40.
 DR N-PSDB; ARO49510.
 XX
 PT High molecular weight surface proteins - of non-typeable
 PT haemophilus which exhibit immunogenic properties
 XX
 XX Claim 5; Figure 10; 100pp; English.
 XX
 CC The isolation and purification of the high molecular weight protein
 CC enables the identification of the major protective epitopes of the
 CC protein by conventional epitope mapping. These epitopes can then be
 CC synthesised using standard techniques and incorporated into fully
 CC synthetic or recombinant vaccines.
 XX
 SQ Sequence 1338 AA;
 Query Match 9.9%; Score 299.5; DB 14; Length 1338;
 Best Local Similarity 22.8%; Pred. No. 6e-11;
 Matches 166; Conservative 98; Mismatches 247; Indels 217; Gaps 36;
 QY 19 TALFTTSTAAML-----TSGCVLG-RARTVTADCAELAGT-NIGPCAGA--FVAGSTLQ 70
 Db 410 TTLNVTSKGFNSIDSTGSGTSPSIRNAELNGITFNKATFNAOGSTANFSAKASIMP 469
 QY 71 Y--TGAFVTVDADVRA-----LDLNNFAAGL-----FSVTGDISLGSVVDTG 112
 Db 470 FKNANYALFENEDISVSGGVSNFKLNASSNIQTGCVIISKQNFNVSGGSTLN--LKA 527
 QY 113 GANKLAVNIDDDGLTLTLTG-----TGT-----AAYGANPALLFQGGQ---AAANNTY 156
 Db 528 GSTETAFSTENDLNLNATGNNITROVEGTSRVNKGVAACKNITFKGNNITFGSOKATT 587
 QY 157 TALGNITLG-GANAGLTIASDPVLGPITLAGNIDGGIITDNDRAINGTGN-----209
 Db 588 EIKGNTYINKNTARGANFAENKSPLNAGNVINNGNL--TAGSIINAGNLTVSKG 645
 QY 210 -----TNPAQI-----SIGASTLSL--GGAVIKATTKLTNAAPVLTITNAN---AV 252
 Db 646 ANLQAITNTYTFNVAGSFDNNGASNISFARGAKFK---DINTSSLTNTNSDYRTI 701
 QY 253 LTCAVDNTTGGDDVGVLN-----LNGALSOVTGNIG-----NTNSLATISVGA---295
 Db 702 IKNISNKG--DLNIIDKSDAEIQIGGNISQEGNLATISSDKVNITNITIKAGVEGG 759
 QY 296 -----GATLGGAVIKATT-309
 Db 760 RSDSSEAEANLTIQKELKLAGDLNLSGNPKAEITAKNGSDLTIGNASGNNADAKRVTF 819
 QY 310 -----TKLTNAASVLTITNNAVLTGAVDNTTGGDNVGVNLSGALSQVNTGNIGTNSLATI 364
 Db 820 DKVKDSKISTDGHNVTLNSEVKTSNGSSNAGNDNSTGLTISA--KDVTVN--NNVTSKHTI 876
 QY 365 NIGA---GVATLDGAVIKATTKTL-----TDDASVLTFTNPVVVTGAIDN-406
 Db 877 NISAAAGNVTTKEGTTINATTSVETQAQNGTIKGNITSONVTATENLVTTENAVINA 936
 QY 407 -----TGANKGVVIFTGAVTNDIGNTAVLAE-----VSVGAG--LLQIOGG 438
 Db 937 TSGTVNISTKTGDIKGIESTSGNVNITAS-GNTLKVSNITGQDVTVTADAGALTITAGS 995
 QY 449 VVKANAINLTDNASVVFTGD--STVTGSIIGGTLPATVNIAGITLIRAGSGSLAANNIDF 506
 Db 996 TISAT-----TGNANITTKTGDKINGKVESSSGVTLVAT-----GATLAVGNI--1038
 QY 507 GAASNLEFNGPAGKNYNLIGTTANGNATNLINAAG-----TVIANDVSIGT 553
 Db 1039 -SGNTVTIITADSGKLTSTVGSTINGTNSVTSSQSGDIEGTISGNTVNTVNTASTGDLTIGN 1097
 QY 554 VAQINIQN 561
 Db 1098 SAKVEAKN 1105

RESULT 14

AAW30291
ID AAW30291 standard; Protein; 1598 AA.

AC AAW30291;

XX 14-APR-1998 (first entry)

XX Non-typeable Haemophilus high mol.wt. surface protein HMW3.

XX Non-typeable Haemophilus; high molecular weight surface protein;
XX HMW3; immunogen; vaccine; otitis media.

XX Haemophilus influenzae strain 5.

XX Key Location/Qualifiers

XX Misc-difference 113

XX /note= "encoded by GTG"

XX Misc-difference 864

XX /note= "encoded by TGT"

XX W09736914-A1.

XX 09-OCT-1997.

XX 01-APR-1997; 97WO-US04707.

XX 01-APR-1996; 96US-0617697.

XX (BARE/) BARENKAMP S J.

XX Barenkamp SJ;

XX WPI; 1997-503038/46.

XX N-PSDB; AAT90992.

XX High molecular weight proteins of non-typeable Haemophilus

XX influenzae - useful for vaccine production

XX Claim 1; Page 93-97; 183pp; English.

XX This protein comprises the high molecular weight surface protein
XX HMW3 (125 kDa) of non-typeable Haemophilus influenzae strain 5 that
XX has the immunological ability to protect against disease caused by
XX a non-typeable Haemophilus strain and is characterised by at least
XX one surface-exposed B-cell epitope that is recognised by monoclonal
XX antibody AD6. The HMW3 amino acid sequence was deduced from an
XX isolated hmw3 gene (see AAT90992). HMW1 (see AAW30293), HMW2 (see
XX AAW30294) and HMW4 (see AAW30292) have also been identified. A
XX conjugate comprising HMW3 linked to an antigen, hapten or
XX polysaccharide, and a synthetic peptide of 6-150 amino acids
XX corresponding to at least protective epitope of HMW3 are also
XX claimed. HMW proteins, conjugates and peptides can be used in
XX vaccines, as immunogens for preparation of antibodies and as
XX antigens for detection of these antibodies.

XX Sequence 1598 AA;

Query Match 9.9%; Score 299.5; DB 18; Length 1598;
Best Local Similarity 22.8%; Pred. No. 7 4e-11;
Matches 166; Conservative 98; Mismatches 247; Indels 217; Gaps 36;

Qy 19 TALFTTSTAAML-----TGSGLV-AARTVADGAELAGT-NIGPGAGA-FVAGSTLQ 70

Db 670 TLANVTSKFNLSIDSTGSGTSPSIRNAELTFNKATFNAGSTANFSIKASIMP 729

Qy 71 Y-TCAFTVTDADYSVRA-----LDLNNFAGL-----FSVTGDISLGSVDYTG 112

Db 730 FKSNNYALFNEDISVSGGSLNFKLNASSNIOTPGVIKSNFNVSGGSTLN--LKAE 787

Qy 113 GANKLAVIDGLTITLTG-----TGT-----AAYGANPALLIFOGGQ---AAANNY 156

Db 788 GSTETAFSIENDLNLNATGGNITITROVEGTSRNVKNVGAACKNITFGKNTFGSQKATT 847

Qy 157 TALGNITLG-GANAGLTITASDPDVLGPITLAGNIDGGIITDNTDAAINCTIGN----- 209
Db 848 EIKGNVTINKNTNATLRGANFAENKSPNLTAGNVINNGNLT--TAGSIINAGNLTVSKG 905
Qy 210 -----TNPAAQI-----SIGASTLSL--GGAVIKATTTKLTAAPVLTLTNAN---AV 252
Db 906 ANLOAITNTYTFNVAGSFDNNGASNISIAKGAKFK----DINNTSSLNITNTSDTYRTI 961
Qy 253 LTGAVDNTTGGDDVGLN-----LNGALSQVGTGIG-----NTNLSATISVGA--- 295
Db 962 IKNISNKSQ--DUNIIDKSDABEIQGGNISQKGNLTISSDKVNTNQTITKAGVEGG 1019
Qy 296 -----GATLGGAVIKATT- 309
Db 1020 RSDSSEAEANLTIQTKELAGDLNISGFNKAETAKNGSDLTIGNASGGSNADAKKVT 1079
Qy 310 -----TKLTNAASVLTLTNAVLTCADVNTTGGDNGVNVNLSGALSQVGTGNTGNITNSLATI 364
Db 1080 DKVKDSKISTDGHNVTLNSEVKTNGSSNAGNDNSTGLITISA--KDVTVN--NNVTSHTKI 1136
Qy 365 NIGA---GVATLDCGAVIKATTTKL-----TDDASVLIFTNPVVVTGAIDN- 406
Db 1137 NISAAAGNVITKEGTTNATGTSVEVTAQNGTIKGNITSQNVTVTATENLVTENAVINA 1196
Qy 407 -----TGNANKGVVIFTGASTVTDNIGNTAVLAE-----VSVGAG--LLQIQGG 448
Db 1197 TSGTVNISTKTGDIKGGIESTSGNVNITAS-GNTLKVSNITGQDVTVTADAGALTTTAGS 1255
Qy 449 VVKANAINLTDNASVVTFTGD--STVTSIGGTGTELFATVNVIGAGITLRAGSGLAANNIDF 506
Db 1256 TISAT-----TGNANITTKGDINGKVESSGSVTLVAT-----GATLAVGNI-- 1298
Qy 507 GAASNLEFNGPAGKNYNLGTFIANGNNATLINAAG-----TVIANDVSIGT 553
Db 1299 -SGNTVITADSGKLTSTVGSTINGTNSVTSSOGDIEGTISGNTVNTASTGDLTIGN 1357
Qy 554 VAQINQN 561
Db 1358 SAKVEAKN 1365

RESULT 15

ABBS2855

ID ABBS2855 standard; Protein; 2732 AA.

XX AC ABBS2855;

DT 11-FEB-2002 (first entry)

XX Escherichia coli polypeptide SEQ ID NO 1119.

XX Escherichia coli; B2/B+A-; anti-inflammatory; antibacterial;
XX immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
XX systemic infection; non-diarrhoeal infection; septicaemia;
XX pyelonephritis; antibiotic resistance.

OS Escherichia coli.

PN WO200166572-A2.

XX 13-SEP-2001.

XX 12-MAR-2001; 2001WO-EP03445.

XX 10-MAR-2000; 2000FR-0003145.

PR 02-FEB-2001; 2001FR-0001449.

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

XX WPI; 2001-550253/61.

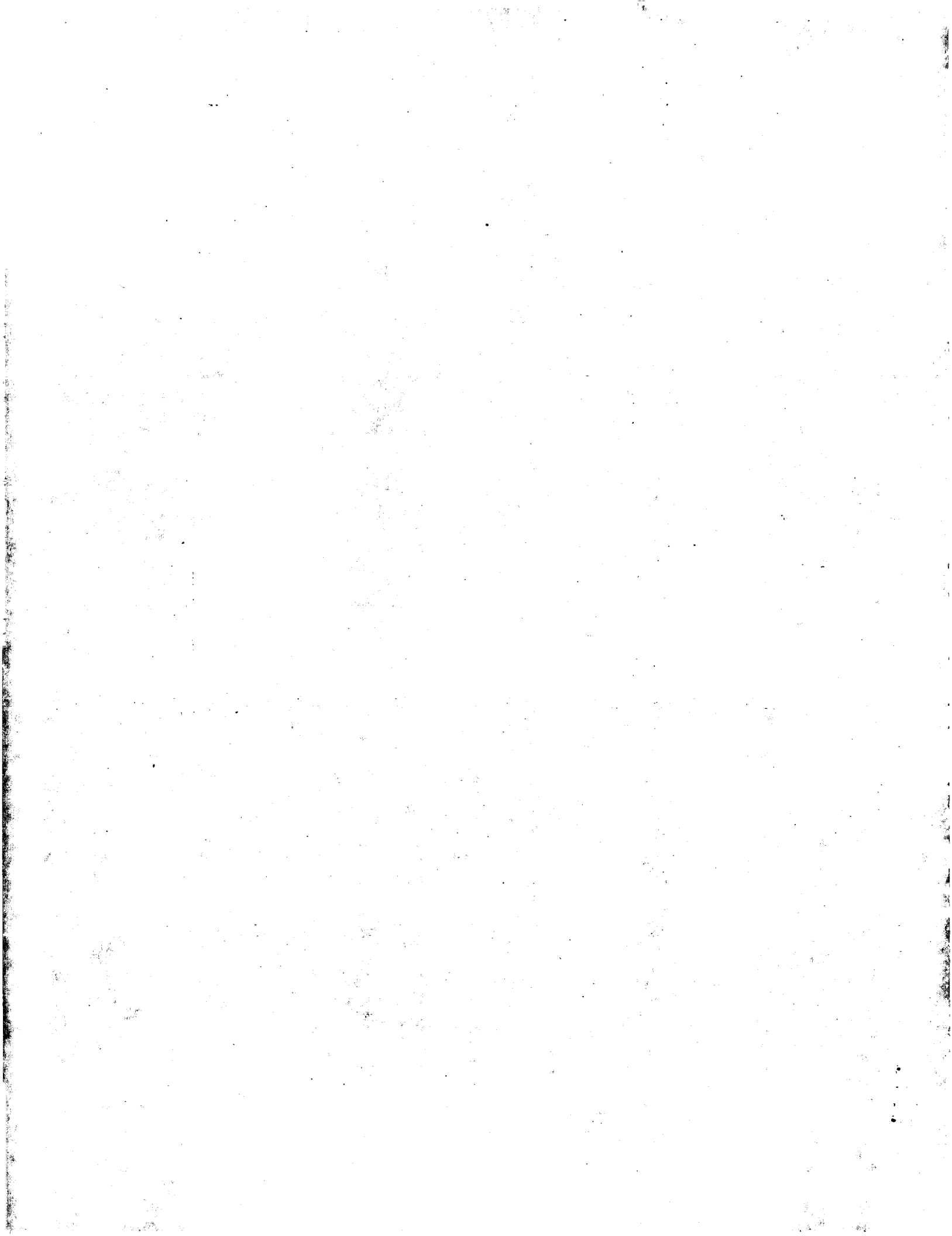
Job time : 53 secs

```

xx  A library of DNA fragments of Escherichia coli strains for the
PT  phylogenetic determination of a given strain comprises polynucleotides of
PT  nature B2/D+ A- -
xx
PS  Example 6; Fig 6; 646pp; English.
xx
CC  The invention relates to a library of DNA fragments of Escherichia coli
CC  strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC  and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC  B2/D+ A- . The polynucleotides have potential antiinflammatory,
CC  antibacterial and immunosuppressive activity as part of pharmaceutical
CC  compositions used to treat, palliate or prevent extra-intestinal E. coli
CC  infections. The polypeptides are useful for determining the phylogenic
CC  group of a given E. coli strain. These polypeptides can detect and treat
CC  an undesired development of E. coli, particularly an extra-intestinal
CC  infection that include systemic and non-diarrhoeal infections such as
CC  septicemia, pyelonephritis and meningitis this is particularly
CC  advantageous as bacterial resistance is increasing with the more
CC  frequent use of broad spectrum antibiotics.
xx
SQ  Sequence 2732 AA;
    Query Match          9.9%; Score 299.5; DB 22; Length 2732;
    Best Local Similarity 25.6%; Pred. No. 1.4e-10;
    Matches 165; Conservative 78; Mismatches 248; Indels 153; Gaps 29;
    QY  24  TSTAAMLTGSGVLGAARTVADGAELAGT-----NIGPGAGAFVA-----G 66
    Db  791 TSTGIMSGDDTLKAA--TTDNRGSLLSAGTSLDGNLSLDNRGTVOGNHVTIRQNSVTNS 849
    QY  67  STLOXTGFTVDADYSVRALDNFAGLFSVTGDISL--GSVVDVTGG----- 113
    Db  850 GTLTGTAALTAAARWASPOPALMNN--GCSLLTSGDLTITAGSITSSGHQWQKRVLTAD 907
    QY  114 --ANKLAVNIDGLTITGTCTAAYGANPALLFOGQQAANNYYTALGNITLGGANAGL 171
    Db  908 SLANSAGIAQADSLATLTG-----ELVSTAGSKVTSNGEMALSALNL--SNSQO 955
    QY  172 TIASDPDV--LGPITLAGNIDGGIITDNTDAINGTIGNTPAAOISIGASTL----- 223
    Db  956 WIAKNLTAKNSLTAGDITG----VDLTITVNTLNQANGKLLSAGVTLKADSVTN 1011
    QY  224 --SLGAVIKATTKLTNA-----APVLTITNANAVLTGAVDNTGGDDVGLNLNGALSO 277
    Db  1012 DGQLQGNVTTITAGQLTNGHILQGETLTLT-----ASGVNRRSG----GVLMSRNALNV 1062
    QY  278 VTGNIGNTNSLATISVGAGTATLGCATLGCATKATTKLTNAASVLTLTNAVLTGAVDNTGGD 337
    Db  1063 STATLSNOSTIOG-----GGVSLNATDRLQNDKILSGSNLTILTAQVLANTGS- 1111
    QY  338 NVGVNLSGALSVQTNIGNTNSLATINIGAV-----ATLDGAVIKATTKLTDDA 389
    Db  1112 --GLVQATLLLDVYVTVNGRVLAT---GSDVKGTTLNNTGTLOGATL---VNYHTFSS 1163
    QY  390 SVLIPTNPVV-----TGADNTGN-----ANKGVVIFTGASTV-----TDNI 427
    Db  1164 GTLLGTSGLVKRGSSLLQNGTGRLYSAGNLLDADQFSQGVQVATGDVTLKIALTLNH 1223
    QY  428 GNTAVLAESVCGALLOQGGVVKANAINLTDNASVVTFTGDSVTGSGTGTELFATVNI 487
    Db  1224 GTLAAGKTLSTSVTNAITNGVYMQGDAMVL---GAGEAFTNNGLTAGK--GNSVFSAGR- 1277
    QY  488 GAGITLRAGGSL--AANNIDFGAASLEFNPAGKNYNLIGTIANGNNATLINAAGTVIA 546
    Db  1278 ---LFLNAPGSLQGGDVLSNRSRDTITSG-----FTGTAGSLTMNVAGTLLN 1322
    QY  547 NDVSTGTVAQINIQNN--KIFVINAKNADVDIILDAQAISFKGAAS 589
    Db  1323 S-----ALIVAGNNKLFTDRLHNQHGDIILAGNSLWVQKDAAS 1359

```

Search completed: April 14, 2003, 16:08:54



GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:08:04 ; Search time 19 seconds

(without alignments)
1991.748 Million cell updates/sec

Title: US-09-800-065-2

Perfect score: 3022

Sequence: 1 MANISLKFQKAIQKGLKTA.....MIELSLKIIYVLLTVVVS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 248812 segs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US09_PUB pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3022	100.0	619	10	US-09-800-065-2
2	331	11.0	1026	10	US-09-379-931-7
3	303.5	10.0	1600	9	US-10-092-880-10
4	299.5	9.9	1599	9	US-10-092-880-9
5	292.5	9.7	3127	9	US-10-184-644-83
6	286	9.5	1477	9	US-10-092-880-4
7	276	9.1	871	10	US-09-886-468-21
8	271.5	9.1	1536	9	US-10-092-880-2
9	271.5	9.0	1819	9	US-10-184-644-39
10	258	8.5	2340	9	US-10-184-644-379
11	256.5	8.5	943	9	US-09-996-634-131
12	256.5	8.5	943	9	US-09-997-181-131
13	256.5	8.5	943	9	US-09-997-182-131
14	254	8.4	1487	9	US-10-184-644-5
15	253.5	8.4	2397	9	US-10-184-644-29
16	251	8.3	2374	9	US-10-184-644-521
17	249	8.2	2039	9	US-10-192-584-7
18	247.5	8.2	1090	9	US-09-759-130B-134
19	247	8.2	1076	9	US-10-028-072-219

20	247	8.2	1076	9	US-10-121-049-219	Sequence 219, App
21	247	8.2	1076	9	US-10-123-904-219	Sequence 219, App
22	247	8.2	1076	9	US-10-140-470-219	Sequence 219, App
23	247	8.2	1076	9	US-10-175-746-219	Sequence 219, App
24	247	8.2	1076	9	US-10-176-918-219	Sequence 219, App
25	247	8.2	1076	9	US-10-176-921-219	Sequence 219, App
26	247	8.2	1076	9	US-10-137-865-219	Sequence 219, App
27	247	8.2	1076	9	US-10-140-474-219	Sequence 219, App
28	247	8.2	1076	9	US-10-142-431-219	Sequence 219, App
29	247	8.2	1076	9	US-10-143-114-219	Sequence 219, App
30	247	8.2	1076	9	US-10-140-002-219	Sequence 219, App
31	247	8.2	1076	9	US-10-142-419-219	Sequence 219, App
32	247	8.2	1076	9	US-10-123-262-219	Sequence 219, App
33	247	8.2	1076	9	US-10-142-423-219	Sequence 219, App
34	247	8.2	1076	9	US-10-121-050-219	Sequence 219, App
35	247	8.2	1076	9	US-10-141-755-219	Sequence 219, App
36	247	8.2	1076	9	US-10-143-032-219	Sequence 219, App
37	247	8.2	1076	9	US-10-184-644-503	Sequence 503, App
38	246.5	8.2	3150	9	US-10-184-644-81	Sequence 81, App1
39	246	8.1	2010	9	US-10-184-644-357	Sequence 357, App
40	244	8.1	1533	9	US-10-184-644-181	Sequence 181, App
41	244	8.1	1786	9	US-10-184-644-481	Sequence 481, App
42	242.5	8.0	1594	9	US-10-184-644-183	Sequence 183, App
43	242.5	8.0	2294	9	US-10-184-644-283	Sequence 283, App
44	239.5	7.9	1723	10	US-09-841-132-394	Sequence 394, App
45	238.5	7.9	1723	10	US-09-841-132-395	Sequence 395, App

ALIGNMENTS

RESULT 1
US-09-800-065-2
; Sequence 2, Application US/098000065
; Patent No. US20020094552A1
; GENERAL INFORMATION:
; APPLICANT: Bouvier, Donald H
; APPLICANT: Crocquet-Valdes, Patricia
; APPLICANT: Stenos, John
; APPLICANT: Walker, David H
; TITLE OF INVENTION: Rickettsia felis Outer Membrane Protein
; FILE REFERENCE: 026.00121
; CURRENT APPLICATION NUMBER: US/09/800,065
; CURRENT FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: US 60/187,323
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Rickettsia felis
.....US-09-800-065-2

Query Match	100.0%	Score	3022	DB	10	Length	619
Best Local Similarity	100.0%	Pred. No.	1.9e-185	Mismatches	0	Indels	0
Matches	619	Conservative	0	Gaps	0		
QY	1	MANISLKFQKAIQKGLKTA	1	MANISLKFQKAIQKGLKTA	1	MANISLKFQKAIQKGLKTA	1
Db	1	MANISLKFQKAIQKGLKTA	1	MANISLKFQKAIQKGLKTA	1	MANISLKFQKAIQKGLKTA	1
QY	61	GAFVAGSTLQYTGAFVTVDADSVSRALDINNFAAGLFSVTGDISLGSVVDVTGGANKLAVN	120				
Db	61	GAFVAGSTLQYTGAFVTVDADSVSRALDINNFAAGLFSVTGDISLGSVVDVTGGANKLAVN	120				
QY	121	IDDLGLTLTGTCTAAYGANPALLFGGQAAANNNTYALGNITLGGANAGLTITASDPDVL	180				
Db	121	IDDLGLTLTGTCTAAYGANPALLFGGQAAANNNTYALGNITLGGANAGLTITASDPDVL	180				
QY	181	GPITLAGNIDGGGIITDNTDAINGTIGNTPAAQISIGASTLSLGGAVIKATTTLTNA	240				
Db	181	GPITLAGNIDGGGIITDNTDAINGTIGNTPAAQISIGASTLSLGGAVIKATTTLTNA	240				

ap 1/25

QY 241 APVLTNANAVLTGAVDNTTGGDDVGVNLNGLALSOVNTGNTGNTSLATISVGAGTATL 300
DB 241 APVLTNANAVLTGAVDNTTGGDDVGVNLNGLALSOVNTGNTGNTSLATISVGAGTATL 300
QY 301 GGAVIKATTTKLTNAASVLTLTNAVLTGAVDNTTGGDDVGVNLNGLALSOVNTGNTGNTS 360
DB 301 GGAVIKATTTKLTNAASVLTLTNAVLTGAVDNTTGGDDVGVNLNGLALSOVNTGNTGNTS 360
QY 361 LATINIGAGVATLDGAVIKATTTKLTDDASVLIPTNPVVVTTGAIDNTGNANKGVVIFTGA 420
DB 361 LATINIGAGVATLDGAVIKATTTKLTDDASVLIPTNPVVVTTGAIDNTGNANKGVVIFTGA 420
QY 421 STVTNIGNTAVLAESVGVAGLLQIQGGVVKANALNTDNASVVTFTGDSVTGSGTGE 480
DB 421 STVTNIGNTAVLAESVGVAGLLQIQGGVVKANALNTDNASVVTFTGDSVTGSGTGE 480
QY 481 LFATVNIAGITLRAGGSLAANNIDFGAASNLEFNGPAGKNYLNLTGANGNATLINA 540
DB 481 LFATVNIAGITLRAGGSLAANNIDFGAASNLEFNGPAGKNYLNLTGANGNATLINA 540
QY 541 ACTVANDVSIGTVAQINQNNKIFVINAKNADVDILDQAISFKGAASRLFLANVSLQM 600
DB 541 ACTVANDVSIGTVAQINQNNKIFVINAKNADVDILDQAISFKGAASRLFLANVSLQM 600
QY 601 IELSLKIIYVLLTVVVS 619
DB 601 IELSLKIIYVLLTVVVS 619

RESULT 2

US-09-379-931-7

; Sequence 7, Application US/09379931

; Patent No. US2002009792A1

; GENERAL INFORMATION:

; APPLICANT: Smith, John

; APPLICANT: Bingle, Wade H.

; APPLICANT: No. US2002009792A1ellini, John F.

; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULOC

; CURRENT APPLICATION NUMBER: US/09/379,931

; CURRENT FILING DATE: 1999-08-24

; PRIOR APPLICATION NUMBER: US 08/614,377

; PRIOR FILING DATE: 1996-03-12

; PRIOR APPLICATION NUMBER: US 08/194,290

; PRIOR FILING DATE: 1994-02-09

; PRIOR APPLICATION NUMBER: US 07/895,367

; PRIOR FILING DATE: 1992-06-09

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7

; LENGTH: 1026

; TYPE: PRT

; ORGANISM: Caulobacter crescentus

US-09-379-931-7

Query Match 11.0%; Score 331; DB 10; Length 1026;
Best Local Similarity 27.3%; Pred. No. 1.3e-13;
Matches 181; Conservative 70; Mismatches 202; Indels 210; Gaps 36;

QY 23 TTSTAATMLTGSVGLGAART-----VTADG-----AELAAAGTNIGPGA 60
DB 413 TTTTGAATVGGTAVTVAQTAGNAVNLTQTQADVTTGNSSTTAVTQTAAAT-----A 467
QY 61 GAFVAGSTLQYTGAFVTDADVS-----VRALDNNPAGL-----FSVVG-- 101
DB 468 GATVAG---RVNGAVTITSAASATPAGTATVTLGSFGAATIDSSALATVNLSTGTS 524
QY 102 -DISLGSVVDTGANKLAVNIDGLTTLTGTTGAAYGANPALLFOGQAAANNVTYALG 160
DB 525 LGIGRGALTATPANTLTNVNG---LTTTGAIT-----DSEAAADGGFT--- 566
QY 161 NITLGGANAGLTIAS-----DPDVLGPITLAGNIDGGG---IITDNTDAAING-TIGNTNPA 213

DB 567 TINIAGSTASSTIASLVAADATTL-----NISGARVTTTSHATAALGTITVNSVGA 619
QY 214 ---AQISIGASTLSLGGG-----VIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGD- 264
DB 620 TLGELATG--LVFTGGAGRDSILGATTKAIVMGAGDDTIVTSSATL-GAGGSVNGDGG 676
QY 265 -DVGVLNGLALSOVNTGNTGNTSLATISVGAGTATLG-----GAVIKATTTK 311
DB 677 TDVLVANVNGSFSADPAFGGFEILRV---AGAAQOSHANGFATLQLOLGATAGATT-- 730
QY 312 LTNA-----SVLTLTNAVLTGAVDNTTGGDNVGVNLNGLALSOVNTGNTSLATINI 366
DB 731 FTVNAVNVGLTVLAAPTGTCTTTLANATGTDVFNLTLSAALAAAGTVA-LAGVETVNI 789
QY 367 GA---GVATLDGAVIKATTTKLTDDASVLIPTNPVVVTTGAIDNTGNANKGVVIFTGAST 422
DB 790 AATDNTTAHVDTTLTQATSAR-----SIVVTG-----NAGLNL----- 823
QY 423 VTDNIGNTAVLAESVGVAGLLQIQGGVVKANALNTDNASVVTFTGDSVTGSGTGE 482
DB 824 -TNTGNTA-----VTSFASAVTGTAPAVTFVSANTTVGEV----- 858
QY 483 ATVNIGAGITLRAGSLAANNIDFGA-ASNLEFNGPAGKNYLNLTGANGNATL-NTNA 540
DB 859 VTIRGGAGADSLTGSATANDTIIGGAGADTLVYTGTT-----DTFTGGTGADIFDINA 911
QY 541 AGT-----VIANDVSIGTVAQINQNNKIFVINAKNADVDILDQAISFKGA-ASRLFLAN 595
DB 912 IGTSTAFVTITDAVG-----DKLDLVGISTNGAIDAGAFGA 949
QY 596 VSL 598
DB 950 VTL 952

RESULT 3

US-10-092-880-10

; Sequence 10, Application US/10092880

; Patent No. US20020164354A1

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J.

; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE

; TITLE OF INVENTION: HAEMOPHILUS

; FILE REFERENCE:

; CURRENT APPLICATION NUMBER: US/10/092,880

; CURRENT FILING DATE: 2002-03-08

; PRIOR APPLICATION NUMBER: 09/155,614

; PRIOR FILING DATE: 1998-09-30

; PRIOR APPLICATION NUMBER: 08/617,697

; PRIOR FILING DATE: 1996-04-01

; PRIOR APPLICATION NUMBER: PCT/US97/04707

; PRIOR FILING DATE: 1997-04-01

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 10

; LENGTH: 1600

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-10-092-880-10

Query Match 10.0%; Score 303.5; DB 9; Length 1600;
Best Local Similarity 24.0%; Pred. No. 1.3e-11;
Matches 163; Conservative 89; Mismatches 249; Indels 177; Gaps 31;

QY 2 ANISLKLFOKAIOKGLKLTALTFTSTAAT-----MLTGS-----GVLGAAATVTTADGAELAA 52
DB 858 ANVTLOADTNSNTGLKKRTLTLGNISVEGNLSLTGANANIVGNLSIAEDSTFKGE----- 913
QY 53 GTNIGPGAGAFVAGSTLQYTGAFVTDADVSVALDINNFAAGLFSVTGDISLGSVVDTG 112
DB 914 -----ASDNLNITGTFT-----NNGTANI-NIKGVVVKLGDINNKG 947

QY 113 GANKLAVNIDDLGLTLTGTGTAAAYGANPALLFOGQAAAANNYYTALGNITLGG---AN 168
 Db 948 GLN-----ITTNASGTOKTLINGNITNEKGLNIRN- IKADREIQIGGNISOKE 995
 QY 169 AGLTIASDP-DVLGPIITAGNIDGGIITDNTDA-----AINGTIGNTPA 213
 Db 996 GNLTISSDKVNITNQITIKAGVEGGR--SDSSEAENANLTIQTKELKLAGDLNISGFNKA 1053
 QY 214 AOISIGASTLSLGA-----VIKATTKLTNAAPVLTLN-----ANAVLTGAVD 258
 Db 1054 ETAKNGSDLTIGNASGGNADAKKVTDFKDKSKISTDGHNVTLNSEVKTSSNGSAGND 1113
 QY 259 NTFG-----GDDVGV-----LNLGALSQVTKNIGNTNSLAT--ISVAGATLTLGA 303
 Db 1114 NSTGLTISKDVVANNVTSHTKTIINISAAAGNVTTKEGTTINATTSVEVTAQNGTKGN 1173
 QY 304 VIKATTKLTNAASVLTTLTNAVLTGAVDNTTGGDNVGVNLSGALSQVTKNIGNTNSLAT 363
 Db 1174 ITSQNVV-TVATENLVTTENAVI-----NATS-----GTVNISTKTDIKGGTESTS--GN 1221
 QY 364 INIGAVATLDGAVIKATTKLTDDASVLIFTPNVPVVTGAIDNTGNANKGVVIFTGASTV 423
 Db 1222 VNITAGNTLVKSNITGQDVTVTADAGALTTAGSTISA---TTGNAN-----I 1267
 QY 424 TONIGNTAVLAEVSVGAGLLQIQ-----GVVKANAINLTDNASVVTFTGDSV----- 472
 Db 1268 TTKTGDIKGVSSGSVTLVATGATLAVGNISGNTVITADSGKLTSTVGSTINGNSV 1327
 QY 473 -----TGSTGGTLPATVNIAG-----ITLRAGGSAAANNIDFGAASNLEFNGPAGKNYL 524
 Db 1328 TTSQSGDIEGTISGNTVNTASTGDLTIGNSAKVEAKN--GAAT---LTAESGLKLTQ 1381
 QY 525 ICTIANGNNATLINA-----AGTVIANDVSIGTV-----AOINTONKIFVINAKN 571
 Db 1382 TGSSTSSNGQTLTAKDSSIAAGNINAANVTLTTLTGTGSKINATSGTL-TINAKD 1440
 QY 572 ADVDLIDQAQISFKGAAS 589
 Db 1441 AKLD-----GAAS 1448

RESULT 4
 US-10-092-880-9
 ; Sequence 9, Application US/10092880
 ; Patent No. US20020164354A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J.
 ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
 ; FILE REFERENCE:
 ; CURRENT APPLICATION NUMBER: US/10/092,880
 ; CURRENT FILING DATE: 2002-03-08
 ; PRIOR APPLICATION NUMBER: 09/155,614
 ; PRIOR FILING DATE: 1998-09-30
 ; PRIOR APPLICATION NUMBER: 08/617,697
 ; PRIOR FILING DATE: 1996-04-01
 ; PRIOR APPLICATION NUMBER: PCT/US97/04707
 ; PRIOR FILING DATE: 1997-04-01
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 9
 ; LENGTH: 1599
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-10-092-880-9
 Query Match 9.9%; Score 299.5; DB 9; Length 1599;
 Best Local Similarity 22.8%; Pred. No. 2.3e-11;
 Matches 166; Conservative 98; Mismatches 247; Indels 217; Gaps 36;
 QY 19 TALFTTSTRAIML-----TSGSVLUG-AARTVTADGAEIAGT-NIGPGAGA--FVAGSTLQ 70
 Db 670 TTLNVTSGKFNLSIDSTGSGTGPSIRNAELNGITFNKATFNIAQSGSTANFSIKASIMP 729

QY 71 Y--TGAFVTVDADSVRA-----LDLNNFAAGL-----PSVTGDISLGSWDTG 112
 Db 730 FKSNNAYALFNEDISVSGGSGVNFKNLASSSNIQTPGVIIKSONENVSGGTLN--LKA 787
 QY 113 GANKLAVNIDDLGLTLTLTG-----TGT-----AAYGANPALLFOGQ-----AAANNYY 156
 Db 788 GSTETAFSIENDLNLNATGGNITIRQVEGTSRNVKNGVAACKNITFKGNITFFGQKATT 847
 QY 157 TALGNITLG-CANAGLTIASDPDVLGPIITAGNIDGGIITDNTDAAINGTIGN----- 209
 Db 848 EIKGNVTTINKNTNATLRGANFAENKSPNLAGNVLNNGNLT--TAGSIIINAGNLTIVSKG 905
 QY 210 -----TNPAAQI-----SIGASTLSL--GGAVIKATTKLTNAAPVLTLNAN-----AV 252
 Db 906 ANLQAITNVTNVAAGSPNNGASNISIRGGAFFK-----DINTSSLNITNDSITYRTI 961
 QY 253 LTGAVDNTTGGDDVGLN-----LNGALSQVTKNIGN-----NTNSLATISVGA--- 295
 Db 962 IKGNISNKG--DLNIIDKSDAEIQIGNISQKEGNTLTSSDKVNITNQITIKAGVEGG 1019
 QY 296 -----GTATLGGAVIKATT-----GTATLGGAVIKATT- 309
 Db 1020 RSDSSEAENANLTIQTKELKLAGDLNISGFNKAETAKNGSDLTIGNASGGNADAKKVT 1079
 QY 310 -----TKLTNAASVLTTLTNAVLTGAVDNTTGGDNVGVNLSGALSQVTKNIGNTNSLATI 364
 Db 1080 DKVKDSKISTDGHNVTLNASEVKTSSNGSAGNDNSTGLTISA--KDVTVN--NNVTSHTKI 1136
 QY 365 NIGA---GVATLDGAVIKATTKL-----TDDASVLIFTPNVPVVTGAIDN- 406
 Db 1137 NISAAAGNVTTKEGTTINATTSVEVTAQNGTKIGNITSONVTVTATENLVTTENAVINA 1196
 QY 407 -----TCNANKGVVIFTGASTVTDNIGNTAVLAE-----VSVGAG-LLQIQGG 448
 Db 1197 TSGTVNISTKTGDIKGGIESTSGNVNITAS--GNTLVKSNITGQDVTVTADAGALTTTAGS 1255
 QY 449 VVKANAINLTDNASVVTFTGD--STVTSIGGTLPATVNIAGITLRAGGSIAANNIDF 506
 Db 1256 TISAT-----TGANITTKGDIKGVSSGSVTLVAT-----GATLAVGNI-- 1298
 QY 507 GAASLEFNGPAGKNYNLIGTIANGNNATPLNINAG-----TVIANDVSIGT 553
 Db 1299 -SGNTVTITADSGKLTSTVGSTINGNTSVTSSQSGDIEGTISGNTVNTASTGDLTIGN 1357
 QY 554 VAQINIQN 561
 Db 1358 SAKVEAKN 1365

RESULT 5
 US-10-184-644-83
 ; Sequence 83, Application US/10184644
 ; Publication No. US2003004930A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3430R1C227
 ; CURRENT APPLICATION NUMBER: US/10/184,644
 ; CURRENT FILING DATE: 2002-06-28
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 83
; LENGTH: 3127
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-83

Query Match 9.7%; Score 292.5; DB 9; Length 3127;
Best Local Similarity 26.2%; Pred. No. 1.4e-10;
Matches 160; Conservative 21; Mismatches 307; Indels 123; Gaps 18;

QY 19 TALFTTTAAIMLTGSGVLGAARTVTADGAELAACTNIG-----PGAGAFVAGSTLOYT 72
DB 1599 TATTTTCTTACCAAGTCATGATATGTAAGTAGAATGATTCGACGAGGAAGTTTAT 1658
QY 73 GAFVTVDADVSRALDLNFAAGLSFVTGDISLSGVSDVTGGANKLAVNIDBLTLT-----128
DB 1659 GAATATGTGTAGTACCTAGTAAAGTGGCAATATTGGCTTATTCCTGCTCTATAG 1716
QY 129 LTGTGTAAYGANPALLFOGGGAANNVTYALGNITLGCANAGLTIASDPVLGPITLAGN 188
DB 1717 TGTGAATGAAGA-----GTAAACAAATTTGTTGACTATTATTA-----AATTATA 1766
QY 189 IDGGIITDNTDAINGT-----IGNTPAAQISIGASTLSLGGAVIKATTTKLTNAA 241
DB 1767 TTAGACCT--TAAGCTGTTTAGCAAGCATTAAACCAATGATGGCTGCTTT--TGAA 1822
QY 242 PVLFLTNANAVLTGAVDNTGGDDVGVNLNG-ALSOVQTNIGNNTNSLATISVGAGTATL 300
DB 1823 ATATTGATG--TGTTGCTGCGAGGATCTGCAAGAACATGTTTATTTTAAATTTA 1880
QY 301 GGAVIKA-----TTKLTNAASVLTITNAVLTVGAVN-----TTGGDNV 339
DB 1881 TAAACAAGTCACTTAATGCGAGTGTCTGAAATCTTATAAGGTTTATACCTTGATAC 1940
QY 340 GVNLGALSQVQTNIGNNTSLATINIGAVATLDGAVIKATTTKLTDDASVLFTNPV 399
DB 1941 GGAATTTACACAGTAGGAGCTTTAGTGAC--ATAGTGT-----1981
QY 400 VTGAIDNTGNANKGVVIFTGASTVTDNIGNTAVLAESVSGAGLQIOGGVVKANALNLD 459
DB 1982 -AGTTTATGGA-----TGGAGGTGCGGTACTAAATGAATGAACGAGTAAATCTT 2033
QY 460 NASV-----VTFTGDSIVTGSIGSTELFAVN-----486
DB 2034 ACTTGGGTAGAGATGGCTTTGCCAACAAAGTGAACGTGTTTGGTGTGTTTAAACTCATG 2093
QY 487 -----IGAGITFLRAGSLAANNIDFGAASNLFEFGPAGKNYNLTGTIANGNNATLNAA 541
DB 2094 AAGTATGGGTTCAGTGGAATGTTTGGAA-CTCTGAAGGATTTAGACAAGCTTTTGAAA 2152
QY 542 GTVIANDVSI 552
DB 2153 GGATAATCATG 2163

RESULT 6

US-10-092-880-4
; Sequence 4, Application US/10092880
; Patent No. US20020164354A1
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/092,880
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 09/155,614
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 08/617,697
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: PCT/US97/04707
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-092-880-4

Query Match 9.5%; Score 286; DB 9; Length 1477;
Best Local Similarity 21.9%; Pred. No. 1.5e-10;
Matches 173; Conservative 89; Mismatches 241; Indels 288; Gaps 36;

QY 42 TTTADGAELAAGTNI---GPGAGAFVAGS-----TLQYTCAFTVT-----78
DB 597 TITGEGKDFRA-NNVSLNGTGKGLNISSVNNLTHNSLGTINISGNTINOTTRKNTSYW 655
QY 79 ---DADVSVALDLN--NF-----AAGL--FSVTGDIS-----104
DB 656 QTSHDHHNVSALNLETGANFTFYKISSNSKGLTQYRSSAGVNFNGVNGNMFNLKEG 715
QY 105 -----LGSVVDVTGGAN-----KLA-VNIDDDG 124
DB 716 AKVNFKLKPNENMNTSKPLPIRFLANITATGGSVFFDIYANHSGRGAELKMSINISNG 775
QY 125 LTLTLTG--TGTAAYGANPALL-----FOGGA--AANNYYTAL---GN 161
DB 776 ANFTLNSHVRGDDAFKINKDLTINATNSFSLRQTKDDFDGYARNAINSTYNISILGN 835
QY 162 ITLGGANAGLTIASD-----PD-----VLGPITLAG-NI 189
DB 836 VTLGGQSSSSITGNITIEKAAVNTLEANNPNOONTDRVIRKLGSLVNGSLTGENA 895
QY 190 DGGIITDNTDAALINGTGN-----NPAAQISIGASTLSLGGAVIKATTTKLTNA 240
DB 896 DIKNLNTSISATPKGTROTTLNITGNTNGTAEINITQGVVKGNTVNDGDLNITHA 955
QY 241 APVLTNANAVLTGAVDNTGGDDV-----GVNLNGALSQVQTNIG-----NTNSL 288
DB 956 K-----RNQRSIIIGDIINKKSLNITDSNNDABEIQGNISQEGNLTITSSDKINITK 1010
QY 289 ATISVG-----AGTATLGGAVIKATTTKLTNAASVLTITNAVLTVG--AVDNTTGGDNV 339
DB 1011 ITKKGIDGEDSSSDATSNANLTKTKELKLTEDLSISGFNKAEITAKDRDLTIGNSND 1070
QY 340 G-----VYNLSGALSQVQTNIG-----356
DB 1071 GNSGAETVTFNNVKDKISADCHNVTLSKVKTSSNGGRESNDTGLTITAKNVE 1130
QY 357 ---NTNSLATINIGAG--VATLDGAVIKAT-----TTKLTDDASVLFTNPVVVTGAID 405
DB 1131 VNKDITSLKTVNTASEKVTVTAGSTINATNGKASITTK-TGDISGTISGNTVSVSATVD 1189
QY 406 NTGNANKGVVIFTGASTVTDNIGNTAVLAESVSGAGLQIOGGVVKANALNLDNASVVT 465
DB 1190 LTTKSGSKIEAKSGEANVTSATGTI-----GGTISGNTVNTANAGDLT 1233
QY 466 FTGDSVTGSGIGGTELFATVNICAGITLRAGSLAANNIDFGAASNLFEFGPAGKNYNLI 525
DB 1234 VNGAEINATEGAATLTAT---GNLTTEAGSSITSTK---QGVDLAONG-----SIA 1281
QY 526 GTIANGNNATLNINAAGTAVTANDVSGTVAQINIQ--NKKIFVINAKNADV-----DILD 578
DB 1282 GSI---NAANVTLTNTGT-----LTTVAGSDIKATSGTLVFINAKDAKLNGDASGDSTE 1331
QY 579 AQATSFKGAAS 589
DB 1332 VNAVNASGSGS 1342

RESULT 7

US-09-886-468-21
; Sequence 21, Application US/09886468
; Patent No. US20020037293A1
; GENERAL INFORMATION:


```
; APPLICANT: Aventis Pasteur Limited
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and us
; FILE REFERENCE: 77813-5
; CURRENT APPLICATION NUMBER: US/09/886,468
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/113,280
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,281
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,282
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,283
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,284
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,285
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,385
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114,050
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,056
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,057
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,058
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,059
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,061
; PRIOR FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 21
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-886-468-21

Query Match          9.1%; Score 276; DB 10; Length 871;
Best Local Similarity 23.1%; Pred. No. 3.6e-10;
Matches 147; Conservative 77; Mismatches 269; Indels 142; Gaps 25;

QY   30 MLTGGVGLGARVTADGAELAGTNI---GPGAGAFVAGSTLOYTGAFVTDADVRA 86
      :|| :::: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|::
Db    8 LLTASALVFSLPHLMAANTDLSSDNYENGSSAAATAKETSDASGT----- 55

QY   87 LDLNFAAGLFVGTGISGV-----VDTCGANKLANVIDDL--TLTTGT 132
      :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|::
Db    56 -----TYLTSDVINVSAITPADKSCFTNTGCALSF-VGADHSLVLQTIALTHD 105

QY   133 GTAYGANPALLFQG-----GQAANNYYTALGNITLGCANAGLTIASDPVLGPITLA 186
      :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|::
Db    106 GAANNNTALSFSGFSLLIDSPATGTSGKGKAICVNTTEGGTATFTDN---ASVTLQ 162

QY   187 GNI---DGGGIITDNTAAINGTIGNTPRAQISIGASTLSLGGAVIKATTTKLTNAAPV 243
      :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|::
Db    163 KNTSEKDGAAVAYSIDL-----KTTTAALLDONTSTKNGCALCSTANTTVCGNSGTV 216

QY   244 LTLTN-----ANAVLTGAVDNT--TGG----DDVGVLNLNCAL--- 275
      :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|::
Db    217 TFSNSTATDKGGIYSKEKDSITLDANTGVVTFKNTAKTGGAWSSDDNALTGMTQVLQF 276

QY   276 -SQVTGNIGNTNS-----LATISVGAGTATLGGAVIKATTTKLTNAASVLILTIN 323
      :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|::
Db    277 ENKTTSAAQANNPEGCGGAICCYLATADTKTGLAISQNQMSETSNTTTTANGGAIYATK 336

QY   324 AVLTGAVDNT--TGGDNVGVVNLSCALSQVDTGNIGNTNSLATINIGAGVATLDGAVIKAT 381
      :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|::
Db    337 CTLDG---NTTLTFDQNTATACGGGAIYETEDES�KSGTCTVTFSTWAKTGALYSKG 393

QY   382 TTKLITDASVLIFNPVVVVTGDAIDNTGN---ANKGVVIIFTGASTVTDNIG--NTAVLAEVS 437
      :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|:: :|::
```

```

Db 1230 SGDGGTSGGVKATKESLTSQNSKIKATTGEANVTSATGTI-----1274
QY 446 QGGVWKANAINETDNASVVF-----TGDSTVGTSGIG-----TELFATVNI 487
Db 1275 -GGTISGTVNVNANAGDLTVNGAEINATECAATLTSSGKLTTAEASHITSAGQVNL 1333
QY 488 -----GAGITLRAGGSLA-----ANNIDFGA-----ASLNFNGPAGKYNLIG 526
Db 1334 SAQDSVAGSINAANVTLLTGTTLTVKGSNINATSGTLVINAKDAELNGAALGNHTVV- 1392
QY 527 TIANGNNATLNTNAGTIVANDVSIGTVQAINQNNKFVINAKNADVDILDQAISPKG 586
Db 1393 -----NAT-NANGSGSVIAT-----TSSRVNITGD-LITINGLN-----IISK 1429
QY 587 AASRLFLANVSLQMIELSLLLKIYPVLLTV 616
Db 1430 GINTVLLKGVKID-----VKYIQPGIASV 1453

```

RESULT 9

```

US-10-184-644-39
; Sequence 39, Application US/10184644
; Publication No. US20030044930A1

```

GENERAL INFORMATION:

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin

```

```

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227

```

```

; CURRENT APPLICATION NUMBER: US/10/184, 644
; CURRENT FILING DATE: 2002-06-28

```

```

; Prior Application removed - See File Wrapper or Palm

```

```

; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 39

```

```

; LENGTH: 1819
; TYPE: DNA

```

```

; ORGANISM: Homo Sapien

```

```

US-10-184-644-39

```

```

Query Match 9.0%; Score 271.5; DB 9; Length 1819;
Best Local Similarity 27.9%; Pred. No. 1.6e-09;
Matches 147; Conservative 16; Mismatches 276; Indels 87; Gaps 19;

```

```

QY 19 TALFTTTSTAALMLTSGVGLGAARVTADGAELEAAGTINIGPGAGFVAGSTLQYTGFTVT 78
Db 1345 TGTATTTCTAGAAAT-AGTTATGTCTTAGGAAATGT-----GGTTAAATTTTGACTTT 1397
QY 79 DADVSVALDLNNPAAGLFSVGTGDISLG--SVVDTGANKLAVN-IDDGLTLT-LTGTGT 134
Db 1398 TA-----CAGGTAAGTGCAGAGGAGAGTGTTCATGAAATGTCTTAATGTAT 1446
QY 135 AAYG-----ANPALLFQGGQA-AANNNTYALGN-ITLGGANAGLTI 173
Db 1447 AATAACATTTACCTTCACGCTCCATCAGATGGAACGAGTTTCAGTAATCAGGAAGTAT 1506
QY 174 ASDPDVLGPTITLAGNIDGGIITDNTDAAINGTIGNTPAAQISIGASTLSLGGAVIKAT 233
Db 1507 AT-----CTATATGATCTTGATATTGTTTATAATAAATTTGAA-----GTCTAAAGACTGCA 1559
QY 234 TTKLTNAAPVLTLTNANAVLTGAVDNTGG--DDVGVLNLNGALSQVTVGNTGNTSLATI 291
Db 1560 TTTTAAACAAGATTAGTATTAATGCGTTGGCCACGAGCAAAAGATAT-----1609

```

```

QY 292 SVGAGTATLGGAVIKATTTKLTNAAASVLTLTNAVLTGAVDNTT-----GGDNVGVNLSGA 347
Db 1610 TTGATTTACTTAAAAATTTTAAATACCGTTTTCATGAAATTTCTCAGTATTTGTAACAGC 1669
QY 348 LSVGTGNTGNVNSLATINIGAGVATLDGAVIKATTTKLTDDASVLIFTNVPVVVTAIDNT 407
Db 1670 AACTTGTCAAAACCTAA--GCATATTTGAATATGATCTCCCATATATTTGAAATTTGAAATC 1726
QY 408 GNANKGV-----VIFTGASTVTDNIGNTAVLAEVSVGAGLLQIOGGVVKANALINTDNASV 463
Db 1727 GTATTGTGTGCTCTGTATATCTGTTAAAAAATTTAAA-----GGACAGAAACCT---TT 1778
QY 464 VTFTGDSVTGTSIGSTELFATVNICAGITLRAGGSLAANNIDFGAA 509
Db 1779 CTTTGTGTATGCATCT-----TTGAATTAAGAAAGTAATGGA 1818

```

RESULT 10

```

US-10-184-644-379

```

```

; Sequence 379, Application US/10184644
; Publication No. US20030044930A1

```

GENERAL INFORMATION:

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin

```

```

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227

```

```

; CURRENT APPLICATION NUMBER: US/10/184, 644
; CURRENT FILING DATE: 2002-06-28

```

```

; Prior Application removed - See File Wrapper or Palm

```

```

; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 379

```

```

; LENGTH: 2340
; TYPE: DNA

```

```

; ORGANISM: Homo Sapien

```

```

US-10-184-644-379

```

```

Query Match 8.5%; Score 258; DB 9; Length 2340;
Best Local Similarity 24.1%; Pred. No. 1.6e-08;
Matches 147; Conservative 18; Mismatches 320; Indels 124; Gaps 16;

```

```

QY 16 GLKTALETTTSTAALMLTSGVGLGAARVTVA-----DGAELAAGTNTGPG-----59
Db 885 GATTATATCTTATGCGCAAAAAGCTTTATATTCCTGGCTTAGGACAGACCAATACCTTA 944
QY 60 --AGAFVAGSTLQYTGFTVTDADVSVALDLNNFAAGLFSVGTGDISLGSVVDTGANKL 117
Db 945 CAATAAAGCTCTACACATTTTC-----AAGGAGTATGCTGGATTTCATGGA---990
QY 118 AVNIDDDLTLTGTGTAAYGANPALLFQGGQAAANNNTYALGNITLGGANAGLTIASDP 177
Db 991 ACTCTAATTCGTACATAAAAAATTTTAAAGTATTTGCTTTGCTTCAGGCAAGTCTGTTTC 1050
QY 178 DVLG-----PITLAGNIDGGIITDNTDAAINGTIGNTPAAQISIGASTLSLG 226
Db 1051 AATGCTGTACTATGTCCTCTAAAGAGAATTTGGTAACCTTGTGATGTTGGTAAGCAGATAG 1110
QY 227 GAVIKATTTKLTNAAAPVLTLTNANAVLTGA---VDNTTGGDDVG---VLNLNGALSQVT 279
Db 1111 GTGAGTTTTGTATAAATCTTTTGTGTTGAGATCAAGCTGAAATGAAACACTGAAAAAC 1170
QY 280 GNTGNTNSLATISVGAGTATLGGAVIKATTTKLTNAAASVLTLTNAVLTGAVDNTTGGDNV 339
Db 1171 ATGGATTTCATT-----TCTATAACACATTTATTTAAGTATATAACACGTTTTTTTGACA 1224

```

QY 340 GVVNLGSLQSVT-----GNIGNTSLATINIGAGVATLDG 375
 DB 1225 AGTGAAGAAITGTTAATCATCTGTCATGTTCTCAATAGATGTAACCTGTAGACTACG 1284
 QY 376 AVIKATTTKLTDDASVLIETPNVVTGAIIDNTG--NANKGVVIFTGASTVTDNIGNTAVL 433
 DB 1285 G-CTATTG-----AAAAATGCTTATTGTAATAATTTCTTATTCGA 1329
 QY 434 AEVSVCAGLLQIQGVVVKANINLTDNASVVTFTGDSVTGSGTGTELFAVNIAGITL 493
 DB 1330 ATATAGAG-----CAGAGAAAGGAATAATATGTTGAAATAAATGTTTGAATC 1379
 QY 494 RAGGSLAANNIDFGAASLEFNCPACKNYNLIG-----TIANGNATLNLINAAGTV 544
 DB 1380 ATGACCCA---AAGAAAT---GTATTGTTGCACTATCCTTCAGAACTGAAGGTT 1431
 QY 545 IANDVSIGT 553
 DB 1432 AATTATTGT 1440

RESULT 11
 US-09-996-634-131
 ; Sequence 131, Application US/09996634
 ; Patent No. US20020172684A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; FILE REFERENCE: 61260
 ; CURRENT APPLICATION NUMBER: US/09/996,634
 ; PRIOR APPLICATION NUMBER: 09/447,135
 ; PRIOR FILING DATE: 2000-01-03
 ; PRIOR FILING DATE: 1997-12-15
 ; PRIOR APPLICATION NUMBER: US 96/10375
 ; PRIOR FILING DATE: 1996-06-14
 ; PRIOR APPLICATION NUMBER: 60/000,254
 ; PRIOR FILING DATE: 1995-06-15
 ; NUMBER OF SEQ ID NOS: 169
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 131
 ; LENGTH: 943
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-09-996-634-131

Query Match 8.5%; Score 256.5; DB 9; Length 943;
 Best Local Similarity 24.6%; Pred. No. 6.9e-09;
 Matches 171; Conservative 79; Mismatches 250; Indels 195; Gaps 35;

QY 10 QKAIQKGLKLTALTTSTAA-----IMLTGSGVLGA----- 39
 DB 66 QAATQRLAAAFEAALAAATVHPAIIISANRALFVSLVSNLLGONAPATAATEAAYEQMW 125
 QY 40 ARTVTA-----DGAELAAGTINIG-----PGAGAFVAGSTLOYT-----GFTVTD 79
 DB 126 AQDAAMFCYHAGASAAVSALTFFGOALPTVAGGGALVSAQAQVTVFRNLGLANVGE 185
 QY 80 ADVSRALDLNINFAAGLFSV--TGDISLGSVWDTGGANKLAVNIDDLGLTLTGTGTAAYG 138
 DB 186 GNVGNG--NVGNFNLGSANIGNTSGNI---GSSNIGFNVGPGCLTAALNNIGFGNTG 240
 QY 139 ANPALLFOGQQAANNYYTALGNITLGGANAGLTIASDPDLVLPITLAGNIDGGIITDN 198
 DB 241 SNN---IGFGNTGSSNN--IGFGNTGDNRGIGLT---GSGLLG-----FGLNSG---TCN 285
 QY 199 TDAANGT-----IGNTPAAQISIGASTLSLGGAVIKATTTKLTNAAPVLTITNANAVLT 254
 DB 286 IGLFNSGTGNGVGNIGS-----GTGNWNGIG-----NSGNSYNT 317

QY 255 GAVDNTTGDDGVVNLNGLNALSQVTCNIGNTSLATISVGAGTATLGGAVIKATTTKLTN 314
 DB 318 GF--GNSGDANTGFFN--SGIANTGVGNAGNYN---TGSYNPGNSNTGGFNMGOYNTGYLN 371
 QY 315 AASVLT-----LTNAVLTGAVDN-----TTGDNVGV 341
 DB 372 SCNYNTGLANSNVNTGAFITGNFNGFLWRGDHQLIFGSPGFFNSTAPSSGFFNSGA 431
 QY 342 VNLGSLQSVTGNIG--NTNSLATINIG---AGVATLDGAVIKATTTKLTDDASVLIETN 396
 DB 432 GSASGFLNSGANNSSGFFNSSGAIGNSLANAGV--LVSGVINSNTVSGLFNMSLVAIT 490
 QY 397 PVVVTG-----AIDNTGNANKGVVIFTGASTVTDNIGNTAFLAEVSVGAG 441
 DB 491 PALISGFFNTGNSMSGFFGPGPVFNGLANRGVNVNIGNA---NIGNVNLGSGNVGDF 546
 QY 442 LLQIQGVVVKANAINLTDNASVVTFTGDSVTGSGIG---GTELFATVNIQA---GITLR 494
 DB 547 NILGSGNL---GSQNLGSGNVGSPNIGS---GNIGFNVGSGSLGNYNIGSGNLGIYNI 600
 QY 495 AGSLAANNIDFGAASLEFN--GPACKNYNLIGTIANGNATLNI-----NAAGTVIA 546
 DB 601 GFNGVDYNVGFCNAG--DFNOGFANTGNNNIGFANTGNN---NIGLGLSDNQOQGFNIA 655
 QY 547 NDVSICT--VAQINIONNKIEVINAKNADVDILDA 579
 DB 656 SCHNSGTGNSGLFNSGTNNVGIFNAGTGNVGTANS 690

RESULT 12
 US-09-997-181-131
 ; Sequence 131, Application US/09997181
 ; Publication No. US20030049269A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; FILE REFERENCE: 61257
 ; CURRENT APPLICATION NUMBER: US/09/997,181
 ; CURRENT FILING DATE: 2001-11-28
 ; PRIOR APPLICATION NUMBER: 09/447,135
 ; PRIOR FILING DATE: 2000-01-03
 ; PRIOR APPLICATION NUMBER: 08/990,823
 ; PRIOR FILING DATE: 1997-12-15
 ; PRIOR APPLICATION NUMBER: US 96/10375
 ; PRIOR FILING DATE: 1996-06-14
 ; PRIOR APPLICATION NUMBER: 60/000,254
 ; PRIOR FILING DATE: 1995-06-15
 ; NUMBER OF SEQ ID NOS: 169
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 131
 ; LENGTH: 943
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-09-997-181-131

Query Match 8.5%; Score 256.5; DB 9; Length 943;
 Best Local Similarity 24.6%; Pred. No. 6.9e-09;
 Matches 171; Conservative 79; Mismatches 250; Indels 195; Gaps 35;

QY 10 QKAIQKGLKLTALTTSTAA-----IMLTGSGVLGA----- 39
 DB 66 QAATQRLAAAFEAALAAATVHPAIIISANRALFVSLVSNLLGONAPATAATEAAYEQMW 125
 QY 40 ARTVTA-----DGAELAAGTINIG-----PGAGAFVAGSTLOYT-----GFTVTD 79
 DB 126 AQDAAMFCYHAGASAAVSALTFFGOALPTVAGGGALVSAQAQVTVFRNLGLANVGE 185
 QY 80 ADVSRALDLNINFAAGLFSV--TGDISLGSVWDTGGANKLAVNIDDLGLTLTGTGTAAYG 138
 DB 186 GNVGNG--NVGNFNLGSANIGNTSGNI---GSSNIGFNVGPGCLTAALNNIGFGNTG 240
 QY 139 ANPALLFOGQQAANNYYTALGNITLGGANAGLTIASDPDLVLPITLAGNIDGGIITDN 198

Db	126	AQDVAAWFGYHAGASAASVALTFFGQALPTPVACGGALVSAAAAQVTVTRFNRNLGLANVGE	188
Qy	80	ADSVRALDLNNFAAGLFSV-TGDISLGSVVDTGANKLAVNTDDGLTLTLTGCTGPAAYG	138
Db	186	GNVGNG--NVGNENLGSANTIGNIGSGNI---GSSNIGFNGVPGLTAAALNNIGFGNTG	240
Qy	139	ANPALLFQGGQAAANTYTTALGNTITGGANAGITIASDPDVLGPIPTTLAGNIDGGGLITDN	198
Db	241	SNN---IGFGNTGSNN--IGFGNTGDNRGIGLT---GSGLLG---FGGLNSG---TCN	285
Qy	199	TDAAINGT---ICNTNPAQISIGASTLSLGGAVIKATTTKLTNAAPVLTLTNANAVLT	254
Db	286	IGLFNSTGTNGVGNCS-----GTGNMGI-----NSGNSYNT	317
Qy	255	GAVDNTTGGDDVGLNLNGALSOVTGNIGNTSLATISVCAGTATLGGAVIKATTTKLTN	314
Db	318	GF--GNSGDANTGFEN-SGIANTGVGNAGYN--TGSYNPGNSNTGGFNQMYNTGYLN	371
Qy	315	AASVLT-----LTNAVLTGAVDN-----TTGGDNVGV	341
Db	372	SGNYNTGLANSGNVNTGAFITGTNFNNGFLWRGDHQGLIFGSPGFFNSTAPSSGFFNSGA	431
Qy	342	VNLGSGALSOVTGNIG--NTNSLATINIG---AGVATLDGAVIKATTTKLTDDASVLIFTN	396
Db	432	GSASGFLUNSGANNSGFFNSGSGAIGNSGLANAGV-LVSGVINSNTVSGFLNMSLVAITTT	490
Qy	397	PVVVTG-----AIDTGNANKGVVIETGASTVTDNIGNTAVLAEVSVGAG	441
Db	491	PALISGFFNTGSNMGSFFGPPFVNLGLANRGVVNIIIGNA---NIGNYNILGSGNVGDF	546
Qy	442	LLQIQGVVVKANILTDNASVVTFTGDSVTVTSIG-----GTFLFATVNIQA---GITUR	494
Db	547	NILGSGNL---GSONILGSGNVGSFNIQS---GNIGVFNVGSGSLGNYNIGSONLGIYNI	600
Qy	495	AGGSLAANNITDFAASNLLEN-CPACKNYNLICTIANGNNATLINI-----NAACTVIA	546
Db	601	GFNGVDYNVGFGNAG---DFNQCFANTGNNGNIGFANTGNH---NICIGLSGDNQQGFNIA	655
Qy	547	NDVSIGT--VAQNIQNKNKIFVNAKNADVDILDA	579
Db	656	SGWNSGTGNSGLNSGTNNYNGIENACTGNVNGIANS	690

nb 601

```

QY      547 NDVSIGT--VAQINIQNNKIFVINAKNADVDILDA 579
      :  ||      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
Db      656 SGNWSTGNSGLNSGTNNVGIENAGTGNVGIAN 690

RESULT 14
US-10-184-644-5
; Sequence 5, Application US/10184644
; Publication NO. US2003004930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P34301C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612

```

```

; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 5
; LENGTH: 1487
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184 -644 -5
Query Match      8.4%; Score 254; DB 9; Length 1487;

```

```

Best Local Similarity 24.5%; Pred. No. 1.7e-08;
Matches 135; Conservative 20; Mismatches 301; Indels 96; Gaps 17;

QY 19 TALFTTAAIMLTGSGVLGAARTVTDAGAEAAAGTNIQPGAGAFVAGSTLOYTGA----74
Db 341 TTTCTTCATGATAAATGCTGTATCCAAATGCTCAGGTGAGAGTGATAGCTA--TGAAGAC 398

QY 75 -----FTVTDADSVRALDLNNFAAGLFSVTGDIISLGSVVDGANKLAVNIDDLGLTL 129
Db 399 GCCTGTTAGGAACAGGCTGCTGAGTTTGGCTTTTCATTTGTTTCATG--TTGATGT 457

QY 130 TGTG-----TAAYGANPALLFOGGAANNVT--ALGNITLGGANAGLTIASDPDVLGP 182
Db 458 TGGGTCACTATTATGCTTCCATGTCATGTTTGGTGCATATGTTACCCAAATACGTA 517

QY 183 I-TLAGNIDGGGLIDNTDAAINGTIGNTPAAQISIGASTLSLGGAVIKAT-TTKLTNA 240
Db 518 TGTATTATCCGGGACT-----AGCTGT-GTTTTT-----CAAAATGCACCTATATTTT 567

QY 241 APVLTLTNANAVLTGAVDNTTGGDDVGVNLNGLALSQVVG-----NIGNTNSLATISVGA 295
Db 568 GCACCTGATCTACAAATTTGGAAGAACCGAAGAGCTATGGACCTGAGATCACTCTTAA 627

QY 296 GTATLGGAVIKATTTKLTNAASVLTLTNAVLTGAVDNTTGGDNVGV-----342
Db 628 GTCACATTTTCTCTT--TGTATATCTGTTGTAGATAGGTTTATCTCTCAGTACA 685

QY 343 -NLGALSQVVG-----NIGNTNSLATISVGA 366
Db 686 CATTCGAATGAGTAGATTTACATTAATGTTTGTCTTTCATTTTATGTTCT 745

QY 367 GAVATLDGAVIKATTTKLTNDASVLITNPVVVTGAIDNTGNANKGVVIFTGASTVDN 426
Db 746 GAGTTTGAATAGTTTATGAAATTTCTTATTTTCATTCATGATAGCTGTAATAT--803

QY 427 IGNVLAESVVGAGLLQIQGGVVKANALINTDNASVVTFTGDSVTGSGIGTFAVYN 486
Db 804 -GTATATAACAAGAC-----TATATGAATGGATAATGATATCAGTT-TTTTATTC 856

QY 487 IGAGITLRAGGS 498
Db 857 TCAGATTTAGAA 868

RESULT 15
US-10-184-644-29
; Sequence 29, Application US/10184644
; Publication NO. US2003004930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P34301C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 29
; LENGTH: 2397
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-29

```

```

Query Match 8.4%; Score 253.5; DB 9; Length 2397;
Best Local Similarity 25.4%; Pred. No. 3.1e-08;
Matches 149; Conservative 15; Mismatches 292; Indels 131; Gaps 20;

QY 19 TALFTTAAIMLTGSGVLGAARTVTDAGAEAAAGTNIQPGAGAFVAGSTLOYTGAFTVT 78
Db 857 TAAATTTTATAGTTATCTTAATAT-TATGATTTTGTATAAAACAGAGAT--TGATCAT 912

QY 79 DADSVRALDLNNFAAGLFSVTGDIISLGSVVDGANKLAVNIDDLGLTLTGTGAAG 138
Db 913 -----TTGTTTGGTTTGAAGTGAAGTCTGACTTTTGTGAATGAT 953

QY 139 ANPALLFOGGAANNVTY-----LGNITLGGANAGLTIASDPDVLGPITLAGNI 189
Db 954 C-----AGGGTTCAGTCTAGATTTGCAATTAATAAATGAAGAGTCTACATTCAGAACATAAA 1007

QY 190 DG-----GGIIT-----DNTDAAINGTIGNTPAAQISIGASTLSLGGAVIKATTTKLTN 239
Db 1008 AGCAGTGTATACAAAGTTTGAATATGATTTAAGCACAGATGATGG-----TTTAAAT 1062

QY 240 AAPVLTLTNANAVLTGAVDNTTGGDDV-----GVNLNGLALSQVVG-----281
Db 1063 AGTTCTCTAATTTTGAATAATCGTCCCAAGCAATAGATTTATGTATATTTGTTTAAATA 1122

QY 282 -----IGNTNSLATISVAGATLGGAVIKATTTKLTNAASVLTLTNAVLTGAVDNTT 335
Db 1123 ATACACCTATTTCAAGTCTGAGTTTGAAG--ATTTACATTTCCCAAGTATTCATTTATG 1180

QY 336 GDNVGVNLSGALSQVVGNTNSLATINIGAV-----ATLDGAVIKATTT 383
Db 1181 AGGTATTTAAGAAGATTA-----TTTGTAGAGAAAAATATTTCTCATTTGATATAAT 1232

QY 384 KLTDDASVLITNPVVVTGAIDNTGNANKGVVIFTGASTVDNIGNTAVLAESVVGAGLL 443
Db 1233 TTT---CTCTGTTTCACTGT--GTGAAAAAAGAGATATTTCCCATTA-----ITLRAG 496

QY 444 IQGGVVKANALINTDNASVVTFTGDSVTGSGIGT-ELFATVNIAG-----ITLRAG 496
Db 1284 GTTGGCCCATTTGCTCAAGAAA--TGTGTATTTCAAGTGACAAATTCGTGCTCTTTTATAG 1342

QY 497 GSLAANNIDFGAASNLFPNGPKGNYNLIGTTANGNNATLINNAAGT 543
Db 1343 GTATATTTCAAAATTT-----TCCTTGTATTTTAGGT 1374

Search completed: April 14, 2003, 16:12:05
Job time : 33 secs

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:08:59 ; Search time 150 Seconds

(without alignments)
2660.599 Million cell updates/sec

Title: US-09-800-065-2

Perfect score: 3022

Sequence: 1 MANISLKFQKAIQKGLKTA.....MIELSLKLIYPVLLTVVVS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644731110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3022	100.0	619	22	US-09-800-065-2
2	1494.5	49.5	4249	4	US-08-029-329-2
3	350	11.6	4327	27	US-60-360-039-10178
4	333.5	11.0	1026	3	US-07-895-367-7
5	333.5	11.0	1026	10	US-08-614-377-7
6	331	11.0	1026	5	US-08-194-290A-7

7	331	11.0	1026	17	US-09-379-931-7	Sequence 7, Appli
8	331	11.0	1026	21	US-09-743-731-5	Sequence 5, Appli
9	323.5	10.7	1242	21	US-09-739-449-13178	Sequence 13178, A
10	323.5	10.7	1242	22	US-09-803-110-13178	Sequence 13178, A
11	322	10.7	1180	15	US-09-167-568-61	Sequence 61, Appli
12	322	10.7	1180	15	US-09-167-568A-61	Sequence 61, Appli
13	322	10.7	1188	15	US-09-167-568-59	Sequence 59, Appli
14	322	10.7	1188	15	US-09-167-568A-59	Sequence 59, Appli
15	316.5	10.5	915	15	US-09-167-568-35	Sequence 35, Appli
16	316.5	10.5	915	15	US-09-167-568A-35	Sequence 35, Appli
17	316.5	10.5	1222	15	US-09-167-568-37	Sequence 37, Appli
18	316.5	10.5	1222	15	US-09-167-568A-37	Sequence 37, Appli
19	316.5	10.5	1228	15	US-09-167-568-34	Sequence 34, Appli
20	316.5	10.5	1228	15	US-09-167-568A-34	Sequence 34, Appli
21	303.5	10.0	1600	15	US-09-155-614-10	Sequence 10, Appli
22	303.5	10.0	1600	24	US-10-092-880-10	Sequence 10, Appli
23	301	10.0	1215	27	US-60-360-039-10073	Sequence 10073, A
24	299.5	9.9	1599	15	US-09-155-614-9	Sequence 9, Appli
25	299.5	9.9	1599	25	US-10-092-880-9	Sequence 9, Appli
26	299.5	9.9	2732	26	US-10-238-075-1119	Sequence 1119, Ap
27	298.5	9.9	2834	24	US-10-085-959-252	Sequence 252, App
28	296	9.8	1060	21	US-09-791-537-65965	Sequence 65965, A
29	296	9.8	1060	21	US-09-791-537-65966	Sequence 65966, A
30	296	9.8	1060	21	US-09-791-537-65967	Sequence 65967, A
31	294	9.7	1060	21	US-09-791-537-65971	Sequence 65971, A
32	293.5	9.7	969	15	US-09-167-568-32	Sequence 32, Appli
33	293.5	9.7	969	15	US-09-167-568A-32	Sequence 32, Appli
34	293.5	9.7	975	15	US-09-167-568-30	Sequence 30, Appli
35	293.5	9.7	975	15	US-09-167-568A-30	Sequence 30, Appli
36	293	9.7	1060	21	US-09-791-537-65972	Sequence 65972, A
37	293	9.7	1465	27	US-60-360-039-13955	Sequence 13955, A
38	292.5	9.7	1054	21	US-09-791-537-103009	Sequence 103009,
39	292.5	9.7	3127	25	US-10-176-912-83	Sequence 83, Appli
40	292.5	9.7	3127	25	US-10-179-524-83	Sequence 83, Appli
41	292.5	9.7	3127	25	US-10-184-634-83	Sequence 83, Appli
42	292.5	9.7	3127	25	US-10-184-644-83	Sequence 83, Appli
43	292	9.7	1612	3	US-07-742-128-2	Sequence 2, Appli
44	289	9.6	1220	15	US-09-167-568-28	Sequence 28, Appli
45	289	9.6	1220	15	US-09-167-568A-28	Sequence 28, Appli

ALIGNMENTS

RESULT 1
US-09-800-065-2
; Sequence 2, Application US/098000065
; GENERAL INFORMATION:
; APPLICANT: Bouyer, Donald H
; APPLICANT: Crocquet-Valdes, Patricia
; APPLICANT: Stenos, John
; APPLICANT: Walker, David H
; TITLE OF INVENTION: Rickettsia felis Outer Membrane Protein
; FILE REFERENCE: 026.00121
; CURRENT APPLICATION NUMBER: US/09/800,065
; CURRENT FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: US 60/187,323
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Rickettsia felis
US-09-800-065-2

Query Match 100.0%; Score 3022; DB 22; Length 619;
Best Local Similarity 100.0%; Pred. No. 5.2e-247;
Matches 619; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MANISLKFQKAIQKGLKLTFTTAAIMLTGSGVLGAARTVTADGAELAGTNTGPGA 60
|||||
DB 1 MANISLKFQKAIQKGLKLTFTTAAIMLTGSGVLGAARTVTADGAELAGTNTGPGA 60
|||||

QY 61 GAFVAGSTLOVGTGFTVTDADVSVRALDNNFAAGLFSVTGDISLGSVVDGAGANKLAVN 120
DB 61 GAFVAGSTLOVGTGFTVTDADVSVRALDNNFAAGLFSVTGDISLGSVVDGAGANKLAVN 120
QY 121 IDDGTLTCTGTAAGANPALLFOGQAAANNNTYALGNITLGGANAGLTIASDPVL 180
DB 121 IDDGTLTCTGTAAGANPALLFOGQAAANNNTYALGNITLGGANAGLTIASDPVL 180
QY 181 GPITLGNIDGGIITDNTDAINGTNTNPAAQISIGASTLSLGGAVIKATTTKLTNA 240
DB 181 GPITLGNIDGGIITDNTDAINGTNTNPAAQISIGASTLSLGGAVIKATTTKLTNA 240
QY 241 APVLTNLNANAVLTGAVDNTGGDDVGVNLNGALSVQVTGNIGNTNSLATISVAGTATL 300
DB 241 APVLTNLNANAVLTGAVDNTGGDDVGVNLNGALSVQVTGNIGNTNSLATISVAGTATL 300
QY 301 GGAVIKATTTKLTNAASVLTLTNAVLGAVDNTGGDNGVNVNLSGALSVQVTGNIGNTNS 360
DB 301 GGAVIKATTTKLTNAASVLTLTNAVLGAVDNTGGDNGVNVNLSGALSVQVTGNIGNTNS 360
QY 361 LATINIGAGVATLGGAVIKATTTKLTDDASVLIETPNPVVVTGAIDNTGNANKGVVIFGA 420
DB 361 LATINIGAGVATLGGAVIKATTTKLTDDASVLIETPNPVVVTGAIDNTGNANKGVVIFGA 420
QY 421 STVTDNIGNTAVLAEVSVGAGLLQIQGVGVKANINLTNDNASVVTFTGDSVTGSGIGTE 480
DB 421 STVTDNIGNTAVLAEVSVGAGLLQIQGVGVKANINLTNDNASVVTFTGDSVTGSGIGTE 480
QY 481 LFAVNTGAGITLRAAGSLAANNIDFGAASNLFEFGAGKNYNLIGTANGNATLINA 540
DB 481 LFAVNTGAGITLRAAGSLAANNIDFGAASNLFEFGAGKNYNLIGTANGNATLINA 540
QY 541 AGTVIANDVSGTVAQINQNNKIFVINAKNADVDIILDAQAISFKGAASRLFLANVSLOM 600
DB 541 AGTVIANDVSGTVAQINQNNKIFVINAKNADVDIILDAQAISFKGAASRLFLANVSLOM 600
QY 601 IELSLKIIIPVLTVVVS 619
DB 601 IELSLKIIIPVLTVVVS 619

RESULT 2

US-08-029-329-2

; Sequence 2, Application US/08029329

; GENERAL INFORMATION:

; APPLICANT: SUMNER W., JOHN

; APPLICANT: ANDERSON E., BURT

; APPLICANT: PELLET E., PHILIP

; APPLICANT: SANCHEZ-MARTINEZ, DEMETRIO

; TITLE OF INVENTION: "BACULOVIRUS EXPRESSED 190 kDa ANTIGENS

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; STREET: 133 Carnegie Way, Suite 400

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: USA

; ZIP: 30303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/029,329

; FILING DATE: 19930309

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: SPRATT, GWENDOLYN D.

; REGISTRATION NUMBER: 36,016

; REFERENCE/DOCKET NUMBER: 1414.049

TELECOMMUNICATION INFORMATION:

; TELEPHONE: 404/688-0770

; TELEFAX: 404/688-9880

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2249 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-029-329-2

Query Match 49.58; Score 1494.5; DB 4; Length 2249;

Best Local Similarity 51.18; Pred. No. 1.1e-116;

Matches 364; Conservative 58; Mismatches 123; Indels 167; Gaps 18;

QY 20 ALETTSTAAILMTGS-----GVL---GAARTVTAD--GAELAAGTNGPQ---A 60
DB 613 SVLTLTANAVLTGAIDNTGGDNGVNLNGALSVQVTGDNIGNTNSLATISVAGTATLG 672
QY 61 GAFVAGSTLOVGTGFTV---TDADVSVRALD--LNNEAAGLFSVTGDISLGSVVDGAGANK 116
DB 673 GAVIKATTTKLTNAASVAVKFTNPVVTGAIDSTGNANNIVTFTGNSVTG--DIGNTNA 730
QY 117 LA-VNIDDLGLTLTGTGTAAAYGANPALLFOGQAAANNNTYALGN-----ITLGGANAGL 171
DB 731 LATVNV-----GAGTATLGG-----AVIKATTTKLTNAASVLTLTNANAVL 771
QY 172 TIASD---PDVLGPITLAGNIDGGIITDNTDAINGTNTNPAAQISIGASTLSLGG 227
DB 772 TGAIDNTGGDNGVNLNGAL-----SQVTGDNIGNTNSLATISVAGTATLGG 820
QY 228 AVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTGGDNGVNLNGALSVQVTGNIGNTNS 287
DB 821 AVIKATTTKLTNAASVLTLTNANAVLTGAVDNTGGDNGVNLNGALSVQVTGDNIGNTNS 880
QY 288 LATSVCAGTATLGGAVIKATTTKLTNAASVLTLT--NAVLTGAVDNTGGDNGVNVNLS 345
DB 881 LATSVCAGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTGGDNGVNLN 940
QY 346 GALSQVTGNIGNTNSLATINIGAGVATLDGAVIKATTTKLTDDASVLIETPNPVVVTGAID 405
DB 941 GALSQVTGDNIGNTNSLATISVAGTATLGGAVIKATTTKLTDAASAVKFTNPVVTGAID 1000
QY 406 NTGNANKGVVIFTGASTVTDNIGNTAVLAEVSVGAGLLQIQGVGVKANINLTNDNAS--- 462
DB 1001 NTGNANKGVVIFTGASTVTDNIGNTAVLAEVSVGAGLLQIQGVGVKANINLTNDNASAVT 1060
QY 463 ----- 462
DB 1061 FTNPVVTGAIDNTGNANNIVTFTGNSVTGNVNTNATLVNAGLQVQGVVYKAN 1120
QY 463 -----VVTFTGDSVTGSGISIGTELFATVNIAG 490
DB 1121 TINLTNDNASAVTFTNPVVTGAIDNTGNANNIVTFTGNSVTGDNIGNTALATVNVGAG 1180
QY 491 ITRAGSGLAANNIDFGAASNLFEFG---AGK---NYNLIGTANGNATININAACTVI 545
DB 1181 ITRAGSGLAANNIDFGAASNLFEFGPLDGGGKAIPYFFKGAINGANNAILNVTK--LLT 1239
QY 546 ANDVSGTVAQINQNNKIFVINAKNADVDIILDAQAISFKGAASRLFLANVS 597
DB 1240 ASHLLTIGTVAEINAGNLTIDASVGDVLTILNAQININFRADSVLVSLSLT 1291

RESULT 3

US-60-360-039-10178

; Sequence 10178, Application US/60360039

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Chen, Xianfeng

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

;; CURRENT APPLICATION NUMBER: US/60/360,039

;; CURRENT FILING DATE: 2002-02-21

;; NUMBER OF SEQ ID NOS: 47374

;; SEQ ID NO 10178

;; LENGTH: 4327

;; TYPE: PRT

;; ORGANISM: magnetite-containing magnetic coccus

;; FEATURE:

;; NAME/KEY: unsure

;; LOCATION: (1)..(4327)

;; OTHER INFORMATION: unsure at all xaa locations

US-60-360-039-10178

Query Match 11.6%; Score 350; DB 27; Length 4327;
Best Local Similarity 25.6%; Pred. No. 3.3e-19;
Matches 169; Conservative 90; Mismatches 236; Indels 166; Gaps 31;

QY 16 GLK--TALFTTSTAALMLTSGVLGAARTV-----TADGAELAACTNIGPGAGAFV 64

Db 3061 GLKDGAAALAAANTS-YITGSVSVYDVATMAQLSAMDTATGLTYGAGLKDSAAALV 3119

QY 65 AGSTLQYTGAFVTDADVS-----VRALD-----LNNFAAGLFSVTGDISLGS---VVDPT 111

Db 3120 ANTNSYVTGAVTVTDAATTAQLGAIDQDTTGTVNYSLAGIKDVTNSNITIDSGNYVANA 3179

QY 112 GGANKLAVNIDGLTLTLTGTAAYGANPALLFQGGQAAA-----NNYTAGLNTTLGG 166

Db 3180 GGA---TITVNDGIANLITDAGTVTGRNVTVDAAASMAQLSQIDNYTTGALKYVTKD 3236

QY 167 ANAGLTIASDPDVLGPITLGNIDGGIITDNTDAAINGT--IGNTNPAQISIGASTLS 224

Db 3237 AVAALVA-----NTNSVTSYAVSVYDAASMAQLSADQD 3272

QY 225 LGGAVIKATTTLTNAAPVLTITNANAVLTGAVDNTGGDDGVNLNGLALSOVGTGNIGN 284

Db 3273 TTGTL---TYTKLTDVAVNL-VTNTNSYVTGSV-NVTVSDIATISQLSSIDANTTGSVTV 3327

QY 285 T---NSLATISVCACT--ATLGGAVIKATT-TKLT-----NAAS 317

Db 3328 TQIGDAAATLATNAGNYVKATHVTVTDAATIAQLFTIDQDNTGSLVYTAGGVKDSAAAN 3387

QY 318 VLTLTNAVLTGAVDNTGGDNGVNLGALSQVGTGNIG-----NTNSLAT 363

Db 3388 LVVNTNSYVTGAV-NVSVTDVTSIAQLSAVDEYTTGTLTYGAGKDSVANLLVNTNSYVT 3446

QY 364 -----INIGAGVATLDGAVIKATT-----TKLTDASVLIFTNPVVVTGAI-----DNT 407

Db 3447 GSYAVSITDVASMANL-SAIQDFTTGTGLNVTKLSDTVSALVANTNSYVTGSVNVITDNA 3505

QY 408 GNANKGVWIFTGASTVTDNIGNITAVLAESVAGLLQIOGGVVKANAIN--LTDNASVVT 465

Db 3506 SWAN-----MSAIDQNTGTLTYTKLSDTAALAAANTNSYVTGSVNVTVTDNATVAQ 3557

QY 466 FTG-DSTVTSIGGTELFATVNIAGITLRAGGSLAANNIDFGAASNLFEFGPAGKNY-- 522

Db 3558 LTTVDAATTGTI---KVASVWDSGNI-----SSNFAYVDGLGVSYVN 3597

QY 523 ---NLIG-TIANGNATLINAGTVIANDVSGTVQAQINQNNKLFIVINAKNADVIDD 578

Db 3598 ANDNVNAVVAQATDATV-----TIADDVLTITDTSNIOGSTFSAVLTAGADATYLD 3651

QY 579 A 579

Db 3652 A 3652

RESULT 4

US-07-895-367-7

;; Sequence 7, Application US/07895367

;; GENERAL INFORMATION:

;; APPLICANT: Smit, John
;; APPLICANT: Bingle, Wade H
;; TITLE OF INVENTION: Bacterial surface protein expression
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Shlesinger, Arkwright & Garvey
;; STREET: 3000 South Eads Street
;; CITY: Arlington
;; STATE: Virginia
;; COUNTRY: USA
;; ZIP: 22202

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/895,367
;; FILING DATE: 19920609
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Garvey, George A
;; REGISTRATION NUMBER: 17737
;; REFERENCE/DOCKET NUMBER: 5946
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-684-5600
;; TELEFAX: 703-836-5288
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1026 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-07-895-367-7

Query Match 11.0%; Score 333.5; DB 3; Length 1026;
Best Local Similarity 24.7%; Pred. No. 1.1e-18;
Matches 181; Conservative 76; Mismatches 232; Indels 243; Gaps 33;

QY 16 GLKTLFTTSTAALMLTSGVL-----CAARTVTAD-CAELAAGT-----NIGP 58

Db 315 GIETMNTSGAAILTNTSSGVTGLTALNTNTSGAATVTAGAGONTATTAQAANNVAV 374

QY 59 GAGAFV-----AGSTLQYTGAFVTDADVSRAALDNNFAAGLFSVTGDISL----- 105

Db 375 DGGANVTASTVGTSGTIT--VCANSAASGTVSVANSSTTTTGAIYVGTGTVAVTQ 432

QY 106 -GSVD-----TGGANKLAVNIDGLTLTLTGTAAYGANPALLFQGGQAAAN 154

Db 433 AGNAVNTLTQADVTVTGNSTTAVTVTQ--TAAATAGATVAGRVNGAVTITDSAAASAT 490

QY 155 TYTALGNITGLGANAGLTIASDPDVLGPITLAG-----NIDG 191

Db 491 TAGKIATVILGSPGAA-TI--DSSALTTVNLSTGTSGLIGRGALPATPTANTLTUNVG 547

QY 192 ---GGIITDNTDAA-----INGTIGNTPAAQISIGASTLSGLGAVIKATTTLTNA 240

Db 548 LTTTGAITDSEAAADGFTTINAGTASSTIASLVAADATTLINISGDARVTTITSTAA 607

QY 241 APVLTITNA-----NAVLTGAVDNT----- 260

Db 608 LTGITVNSVGATFLGAELATGLVFTGGAGADSILLGATTKAIYMGAGDDTVTVSSATLGA 667

QY 261 ---TGSD--DVGVLNGLALSQVGTGNIGNTNSLATISVCAGTATLG-----G 302

Db 668 GGSVNGGDDVLVANVNGSSFSADPAFGFETLR-----AGAAOGSHNANGFTALQLG 723

QY 303 AVIKATTTTLTNA-----SVLTLTNAVLTGAVDNTTGGDNGVNLGALSQVGTGNIG 357

Db 724 ATAGATT--FTNVAVNVGLTVLAAPTCTTIVTLANATGTSDFENLTSSSAAALAGTVA- 780

QY 358 TNSLATINIGA---GVATLDGAVIKATTTKLTDASVLIFTNPVVVTVTAIDNTGNANKG 413

Db 781 LAGVETVNIATDNTTAHVDTLTLQATSAK-----SIVVTG-----NAG 820
 Qy 414 VVIETGASTVDNIGNTAVLAESVAGLLQIOGVVKANAINLTNDASVVVTFGDSVT 473
 Db 821 LNL-----TNTGNTA-----VTSFASAVTGTCSAVTFVSANTTV 855
 Qy 474 GSIGGTELFATVNIAGITLRAGSLAANNIDFGA-ASNLEFNGPAGKNYNLIGTIANGN 532
 Db 856 GEV-----VTIRGAGADSLTGSATANDTIIGGAGADTLVYTGTT-----DTFTGGT 902
 Qy 533 NATL-NINAAGT-----VIANDVSIGTVAQINQNNKIFVINAKNADVDILLDAQAISFKGA 587
 Db 903 GADIFDINAIGTSTAFVITIDRAVG-----DKLDLVGISTNGA 940
 Qy 588 -ASRLFLANVSL 598
 Db 941 IADGAFGAAVTL 952

RESULT 5

US-08-614-377-7
 ; Sequence 7, Application US/08614377
 ; GENERAL INFORMATION:
 ; APPLICANT: Smit, John
 ; APPLICANT: Bingle, Wade H.
 ; APPLICANT: Nongellini, John F.
 ; TITLE OF INVENTION: EXPRESSION AND SECRETION OF
 ; TITLE OF INVENTION: HETEROLOGOUS
 ; TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson PC
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/614,377
 ; FILING DATE: 12-MAR-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/194,290
 ; FILING DATE: 09-FEB-1994
 ; CLASSIFICATION: 435
 ; APPLICATION NUMBER: US 07/895,367
 ; FILING DATE: 09-JUNE-1992
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Tsao, Y. Rocky
 ; REGISTRATION NUMBER: 34053
 ; REFERENCE/DOCKET NUMBER: 08106/002001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-542-5070
 ; TELEFAX: 617-542-8906
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1026 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-614-377-7

Query Match 11.0%; Score 333.5; DB 10; Length 1026;
 Best Local Similarity 24.7%; Pred. No. 1.1e-18;
 Matches 181; Conservative 76; Mismatches 232; Indels 243; Gaps 33;

Qy 16 GLKTALEFTTAAIIMLTGSGVL-----GAARTVTAD-GAELAAAGT-----NITGP 58
 Db 315 GIETMNVTSAAIILNTSSGVTGLTALNTNTSGAAQVTTAGAGQNLTAATAQAANNVAV 374
 Qy 59 GAGAFV-----AGSTLQYTGAFVTDADYSVRALDLNFAAGLFSVVTGDISL----- 105
 Db 375 DGGANVTIVASTGVTSCTTT--VGANSAASGTVSVSVANSSTTTTCAIAVTCGTAVTVTAQT 432
 Qy 106 -GSVVD-----TGGANKLAVNIDDDGLTLTLTGTTGTAAYGANPALLFGGQGAANN 154
 Db 433 AGNAVNTLTOADVTVTGNSTTAVTVTQ--TAATAGATVAGRVNGAVTITDASAASAT 490
 Qy 155 TYTALGNITLGGANAGLTIASDPDVLGPITLAG-----NIDG 191
 Db 491 TAGKIATVTLGSPCAA-TI--DSSALTTVNLSGTSLGIGRGALTATPTANTLTLNVNG 547
 Qy 192 ---CGIITDNTDAA-----INGTGNTPAAQISIGASTLSLGGAVIKATTTKLTNA 240
 Db 548 LTTTGAITDSEAAADGFTTINIAGSTASSTIASLVAADATTLNISGDARVTTISHTAAA 607
 Qy 241 APVLTLTNA-----NAVLTGAVDNT----- 260
 Db 508 LTGITVINSVGATLGAELATGLVFTGGAGADSILLGATTKAIVMGAGDDTVTVSSATLGA 667
 Qy 261 ---TGGD--DVGVLNNGALSQVGTGNIGNTNLSLTIATSVGAGTATLG-----G 302
 Db 668 GGSVNGGDTDLVANVNGSSFSADPAFGFPETLRV---AGAAAQSGSHNANGFTALQLG 723
 Qy 303 AVIKATTTKLTNAA-----SVLTLTNAVLTGAVDNTTGGDNVGVNLSGALSQVGTGNIGN 357
 Db 724 ATAGATT--FTNVAVNVLTVLAAPTGTTVTLANATGTSDFNLTLSSSALAAGTVA- 780
 Qy 358 TNSLATINIGA---GVATLDGAVIKATTTKLTDDASVLIFTPVWPVVGATDNTGNANKG 413
 Db 781 LAGVETVNIATDNTTAHVDTLTLQATSAK-----SIVVTG-----NAG 820
 Qy 414 VVIETGASTVDNIGNTAVLAESVAGLLQIOGVVKANAINLTNDASVVVTFGDSVT 473
 Db 821 LNL-----TNTGNTA-----VTSFASAVTGTCSAVTFVSANTTV 855
 Qy 474 GSIGGTELFATVNIAGITLRAGSLAANNIDFGA-ASNLEFNGPAGKNYNLIGTIANGN 532
 Db 856 GEV-----VTIRGAGADSLTGSATANDTIIGGAGADTLVYTGTT-----DTFTGGT 902
 Qy 533 NATL-NINAAGT-----VIANDVSIGTVAQINQNNKIFVINAKNADVDILLDAQAISFKGA 587
 Db 903 GADIFDINAIGTSTAFVITIDRAVG-----DKLDLVGISTNGA 940
 Qy 588 -ASRLFLANVSL 598
 Db 941 IADGAFGAAVTL 952

RESULT 6

US-08-194-290A-7
 ; Sequence 7, Application US/08194290A
 ; GENERAL INFORMATION:
 ; APPLICANT: Smit, John
 ; APPLICANT: Bingle, Wade H.
 ; TITLE OF INVENTION: BACTERIAL SURFACE PROTEIN EXPRESSION
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson PC
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: US
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows95
 ; SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194,290A
FILING DATE: 19-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/895,367
FILING DATE: 09-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Tsao, Y. Rocky
REGISTRATION NUMBER: 34,053
REFERENCE/DOCKET NUMBER: 08106/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1026 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-194-290A-7

Query Match 11.0%; Score 331; DB 5; Length 1026;
Best Local Similarity 27.3%; Pred. No. 1.8e-18;
Matches 181; Conservative 70; Mismatches 202; Indels 210; Gaps 36;

```
QY 23 TTSTAAILMTGSSVLGAART-----VTADG-----AELAAAGTNIQPGA 60
DB 413 TTTTGAIVTGGTAVTVAQTAGNAVNTLTQADVTVTGNSSTAVTQTAAAT-----A 467

QY 61 GAFVAGSTLQYTGAFVTDDVYS-----VRALDINFAAGL-----FSVTG-- 101
DB 468 GATVAG---RVNGAVITDSSAASATTAGKIATVILGSFGAATIDSSALTIVNLSGTGFS 524

QY 102 -DISGVSVDGTGANKLAVNIDDLTLTGCTGTAAYGANPALLFGGGQAANNVYTAG 160
DB 525 LGIGRGALTATPTANTLTNLVNG---LTTTGAIT-----DSEAAADDDGFT--- 566

QY 161 NITLGGANAGLTAS----DPDVLGPITLAGNIDGG--IITDNTDAAING-TIGNTPA 213
DB 567 TINIAGSTASSIASLVAADATTL-----NISGDARVTIISHTAAALTGITVTVNSVGA 619

QY 214 ---AQISIGASTLSLGA-----VIKATTTKLTNAAPVLTNANAVLTGAVDNTTGGD- 264
DB 620 TLGAELATG--LVFTGGAGRDSILLGATTKAIVMGAGDDTVTVSSATL-GAGGSVNGDG 676

QY 265 -DVGVLNMGALSQVGTGNIGTNSLATISVGAGTATLG-----GAVIKATTTK 311
DB 677 TDVLVANVNGSFSADPAFGFETLRV----AGAAAGSHNANGFTALQIGATAGATT-- 730

QY 312 LTNA-----SVTLTLNAVLTGAVDNTTGGDNVGVNLSGALSQVGTGNIGTNSLATINI 366
DB 731 FTVNAVNVGLTVLAAPTGTITVTLANATGSDVFNLTLSAALAAGTVA-LAGVETVNI 789

QY 367 GA----GVATLDGAVIKATTTKLTDDASVLIETNPVVVTVTGADNTGNANKGVVIFTGAST 422
DB 790 AATDTNTTAHVDTLTQATSAK-----SIVVTG-----NAGLNL----- 823

QY 423 VTDNIGNTAVLAESVVGAGLLQIGGVVKANALINDNASVVTFTGDSVTGSGIGTELF 482
DB 824 --TNTGNTA-----VTSFASAVTGTAPAVTVFSANTTVGEV----- 858

QY 483 ATVNICAGITLRAGGSIAANNIDFGA-ASNLEPENGAGKNYNLIGTIANGNATL-NINA 540
DB 859 VTIIRGAGADSLTGSATANDTIIGGAGADTLVYTGTT-----DFTTGTGADIPIEDINA 911

QY 541 AGT-----VIANDVSGITVAQINQNNKIFVINAKNADVDILDAQAISFKGA-ASRFLAN 595
DB 912 IGTSTAFVITDAVG-----DKLDLVGISTNGAIDAGFGAA 949

QY 596 VSL 598
|:|
```

Db 950 VTL 952

RESULT 7

US-09-379-931-7
Sequence 7, Application US/09379931
GENERAL INFORMATION:
APPLICANT: Smit, John
APPLICANT: Bingle, Wade H.
APPLICANT: Nomellini, John F.
TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CA
FILE REFERENCE: 08106/002003
CURRENT APPLICATION NUMBER: US/09/379,931
CURRENT FILING DATE: 1999-08-24
PRIOR APPLICATION NUMBER: US 08/614,377
PRIOR FILING DATE: 1996-03-12
PRIOR APPLICATION NUMBER: US 08/194,290
PRIOR FILING DATE: 1994-02-09
PRIOR APPLICATION NUMBER: US 07/895,367
PRIOR FILING DATE: 1992-06-09
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 1026
TYPE: PRT
ORGANISM: Caulobacter crescentus
US-09-379-931-7

Query Match 11.0%; Score 331; DB 17; Length 1026;
Best Local Similarity 27.3%; Pred. No. 1.8e-18;
Matches 181; Conservative 70; Mismatches 202; Indels 210; Gaps 36;

```
QY 23 TTSTAAILMTGSSVLGAART-----VTADG-----AELAAAGTNIQPGA 60
DB 413 TTTTGAIVTGGTAVTVAQTAGNAVNTLTQADVTVTGNSSTAVTQTAAAT-----A 467

QY 61 GAFVAGSTLQYTGAFVTDDVYS-----VRALDINFAAGL-----FSVTG-- 101
DB 468 GATVAG---RVNGAVITDSSAASATTAGKIATVILGSFGAATIDSSALTIVNLSGTGFS 524

QY 102 -DISGVSVDGTGANKLAVNIDDLTLTGCTGTAAYGANPALLFGGGQAANNVYTAG 160
DB 525 LGIGRGALTATPTANTLTNLVNG---LTTTGAIT-----DSEAAADDDGFT--- 566

QY 161 NITLGGANAGLTAS----DPDVLGPITLAGNIDGG--IITDNTDAAING-TIGNTPA 213
DB 567 TINIAGSTASSIASLVAADATTL-----NISGDARVTIISHTAAALTGITVTVNSVGA 619

QY 214 ---AQISIGASTLSLGA-----VIKATTTKLTNAAPVLTNANAVLTGAVDNTTGGD- 264
DB 620 TLGAELATG--LVFTGGAGRDSILLGATTKAIVMGAGDDTVTVSSATL-GAGGSVNGDG 676

QY 265 -DVGVLNMGALSQVGTGNIGTNSLATISVGAGTATLG-----GAVIKATTTK 311
DB 677 TDVLVANVNGSFSADPAFGFETLRV----AGAAAGSHNANGFTALQIGATAGATT-- 730

QY 312 LTNA-----SVTLTLNAVLTGAVDNTTGGDNVGVNLSGALSQVGTGNIGTNSLATINI 366
DB 731 FTVNAVNVGLTVLAAPTGTITVTLANATGSDVFNLTLSAALAAGTVA-LAGVETVNI 789

QY 367 GA----GVATLDGAVIKATTTKLTDDASVLIETNPVVVTVTGADNTGNANKGVVIFTGAST 422
DB 790 AATDTNTTAHVDTLTQATSAK-----SIVVTG-----NAGLNL----- 823

QY 423 VTDNIGNTAVLAESVVGAGLLQIGGVVKANALINDNASVVTFTGDSVTGSGIGTELF 482
DB 824 --TNTGNTA-----VTSFASAVTGTAPAVTVFSANTTVGEV----- 858

QY 483 ATVNICAGITLRAGGSIAANNIDFGA-ASNLEPENGAGKNYNLIGTIANGNATL-NINA 540
DB 859 VTIIRGAGADSLTGSATANDTIIGGAGADTLVYTGTT-----DFTTGTGADIPIEDINA 911

QY 541 AGT-----VIANDVSGITVAQINQNNKIFVINAKNADVDILDAQAISFKGA-ASRFLAN 595
|:|
```

```

Db      912  IGTTAFVTTDAAG-----DKLDLVIGISTNGATADGAFGAA  949
Qy      596  VSL 598
        !:!
Db      950  VTL 952

RESULT 8
US-09-743-731-5
; Sequence 5, Application US/09743731
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; TITLE OF INVENTION: CLEAVAGE OF CAULOBACTER PRODUCED
; TITLE OF INVENTION: RECOMBINANT FUSION PROTEINS
; FILE REFERENCE: 08106-004001
; CURRENT APPLICATION NUMBER: US/09/743,731
; CURRENT FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: PCT/CA99/00637
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: CA 2,237,704
; PRIOR FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 5
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-743-731-5

```

Query Match	11.0%	Score 331;	DB 21;	Length 1026;
Best Local Similarity	27.3%;	pred. No. 1.8e-18;		
Matches	181;	Conservative	70;	Mismatches 202; Indels 210; Gaps 36;
QY	23	TTSTAAMLTGSGVLGAART-----VTAGD-----AELAAGNIGPGA	60	
		: : :		
Db	413	TTTTGALAVTGGTAATVQAQTAGNAVNNTLTQAQVTVTGNSSTTAVTQTAAAA--	A	467
		: : : :		
QY	61	GAFVAGTLQVTAFTTADADV-----VRDLNLFNAAGL-----FSVTG--	101	
		: : : :		
Db	468	GATVAG---RVNGAVTITDSAAAATTAGKATATVLSFGAATIDSSALTIVNLSTGTS	524	
		: : : :		
QY	102	-DISLGSVDVTGGANKLAVIDDGLTTLTGCTGTAAVCANPALLFQGQOAAANTYTALG	160	
		: : : : : :		
Db	525	LGIKGALATPTAPTNTLTLNVNG-----LTTTGAIT-----DSEAAADGFT---	566	
		: : : : : :		
QY	161	NITLGGANAGLTIAS----DPDVLGPITLAGNIDGG--IITDNTDAING-TIGNMPA	213	
		: :		
Db	567	TINIAGSTASSTIASVAADATTL-----NISGDARVITSHTAALGITVTVNSGA	619	
		: :		
QY	214	---AQISIGASTLSLGA-----VIKATTKLTNAAPVLTLTNANAVLTGAVDNNTGGD-	264	
		: : : : : :		
Db	620	TLGAELATG--LVFTGGAGRDSILLGATTKAIYMGAGDDTVTVSSATL-GAGGSVNGDG	676	
		: : : : : :		
QY	265	-DVGVLNGLALSQVTGNIGNTNSLATISVCGATATLG-----CAVLIKATTK	311	
		: : : : : :		
Db	677	TDVLVANVNGSSPADPAFGGFETLRV----AGAAAGSHNANGFTALQLGATAGATT--	730	
		: : : : : :		
QY	312	LTNAA-----SVLTLTNAVLTGAVDNNTGGDNVGVNLGALSQVTGNIGNTNSLATINI	366	
		: : : : : :		
Db	731	FTNVAVNVGLTVLAAPTGTGTTVTLANATGTSDVFENLTLSSSAALAAGTVA-LAGVEIVNI	789	
		: : : : : :		
QY	367	GA---GVATLDGANIKATTKTLTDDASVLIFNPVVVTCIDNTGNANGKVIFTGAST	422	
		: : : : : :		
Db	790	AATDTNTTAHVDFILTLQATSAK-----SIVVTG-----NAGLNL-----	823	
		: : : : : :		
QY	423	VTDNIGNTVAELVSVSGAGLLQGGVYVKANAINLTDNASWVTFITGDSVTGTSIGGTELF	482	
		: : : :		
Db	824	--TNTGNTA-----VTSFDASAVTCTAPATFVSANTTVGEV-----	858	
		: : : : : :		
QY	483	ATVNIAGITLRAGGSAAANNIDFGA-ASNLEFNPGPKNYNLIGTIANGNNATL-NINA	540	
		: : : : : :		
Db	859	VTIRGAGADSLGTSATANDITGGAGADPLVTVTGT-----DFTGTGTGADIFDNA	911	
		: : : : : :		

```

Qy 541 AGT-----VIANDYSIGTVAQIONTNNKIFVINAKNADVDILDQAQATLSFKGA-ASRLFLAN 595
      || | | :||
Db 912 IGTTAFVTTDAVG-----DKDLVLGVGTNGCAITADGAFGAA 949
      || | | :||

Qy 596 VSL 598
      |||
Db 950 VTL 952

RESULT 9
US-09-739-449-13178
; Sequence 13178, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 13178
; LENGTH: 1242
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-09-739-449-13178

```

Query Match	10.7%;	Score 323.5;	DB 21;	Length 1242;
Best Local Similarity	25.0%;	Pred. No. 1e-17;		
Matches 163;	Conservative 90;	Mismatches 195;	Indels 205;	Gaps
QY	20	ALFTTSTAATMLTSGVLG----	AARTVADGAEALAG--TNIG-----	PG-----59
Db	233	AKLVEEAATVTGSEALIGHRSASEATVTGDSGSKWTGDLQVGGDTSPGGAGNGTLNV	292	
QY	60	-AGAFVAGSTLQYTGAF-----	TVTDADVSVRALDLNPF-----	AAGLSFVSTGDISLGS 107
Db	293	TAGGSV--DSTVAHLGVVAGATGSAIVDGRKGSVMTVDRNSLEVSGSGSLAVTG----	GG 347	
QY	108	VVD-----	TGGANKLAVNIDDLGLTLTGCTAAYGANPALLFQGGQAAANNITYA	158
Db	348	LVDAAANIITNTGGNGSVRVSGADSTVKRSDLNVGLYNGSGMTVEAGGVKSRDGYVA	407	
QY	159	LGNITLGGANAGLTIASDPDLVLPITLAGNIDGGIITDNTDAAINGT--	IGNTPAAQI 216	
Db	408	-----	TGGSTSAVTVG-----	DGSSKAMTGTFFVCYASGAT-- 440
QY	217	SIGASTLSLGGV--	IKATTTKLTNAAPVLTLTNAVALTGAVDNTT-----	GGDDVGVNL 271
Db	441	--GNVTVSNGGAIRATGVTLGLDLAGASGTMTITGAGSKVTAYVDINGTVNSGSDVDFGQ-	497	
QY	272	NGALSOVTGN-----	IGNTNSLTIATISVG-----	AGTATL--GGA 303
Db	498	SGSLSVVNGSLDAYNLYVGNALGSSGAVLVSGVSHSVSDGLMVVGNAGNGSVBITGGA	557	
QY	304	VIKATTTKL--TNAASVLTLT-----	NAVLTCAYDNTT--GGDNVGVV-----	342
Db	558	SLAAPTILTATEAGSTGVLSIGAGSGQTARSAGAVEARAIAFAGAGNGSVFNHSETGYTL	617	
QY	343	--NLSCALSOVTCNIGNTNSLATINIGAVATLDCAVIKATTK-----	LTDASVLI-- 393	
Db	618	SADISGA--GRVVAEAGVT--TLSGNNSYSGGTTISAGMLKGTAKSEFGSGGIVNNAELVVDG	675	
QY	394	---FTNPVVVVTGAIDNTGNANKGVVIFTGASTVTDNIGNTAVLAESVSGVAGLLQIQGGW	450	
Db	676	GGTFLSNAISGTSFEKGTGDGN--	LLLLTGNSTYS--GATA-----	VSAGKLSVNGSL- 722
QY	451	KANAINLTNASVVTFTGDSTV--	TGSGTGGTELFATVNIAGITURPAGGSLAANNIDFGA	508
Db	723	-----	ASAVSGSGATVGGTGTIG-----	GLTVNSGGLTAPGN--s 756

QY 509 ASNLEFNGPAGKNYLNIGTIANNNATLINAAGT-----VIANDVSI 551
 Db 757 IGTLTSTGNA-----TFASGSYAVEIDADGSSDRLAVTGTITIANDVSL 801

RESULT 10
 US-09-803-110-13178
 ; Sequence 13178, Application US/09803110
 ; GENERAL INFORMATION:
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
 ; FILE REFERENCE: 38-10(15490)D
 ; CURRENT APPLICATION NUMBER: US/09/803,110
 ; CURRENT FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 09/739,449
 ; PRIOR FILING DATE: 2000-12-19
 ; PRIOR APPLICATION NUMBER: US 09/514,000
 ; PRIOR FILING DATE: 2000-02-23
 ; PRIOR APPLICATION NUMBER: US 60/168,139
 ; PRIOR FILING DATE: 1999-12-01
 ; NUMBER OF SEQ ID NOS: 13351
 ; SEQ ID NO 13178
 ; LENGTH: 1242
 ; TYPE: PRK
 ; ORGANISM: Agrobacterium tumefaciens
 ; US-09-803-110-13178

Query Match 10.7%; Score 323.5; DB 22; Length 1242;
 Best Local Similarity 25.0%; Pred. No. 1e-17;
 Matches 163; Conservative 90; Mismatches 195; Indels 205; Gaps 36;
 QY 20 ALFTTSTAALMLTSGVVG---AARTVTADGAELAAAG-TNIG-----PG-----59
 Db 233 AKLVPEAAATVTSGEALIGHRSASEATVTGDSKWTGTGDLQGGTSDPGGLAGNGLTNV 292
 QY 60 -AGAFVAGSTLQYTGAF-----TVTDADVSRLDLNPF-----AAGLFSVTGDISLGS 107
 Db 293 TAGGSV-DSTVAHLGVVAGATGSAIVDGGKSVWTVDRNSLEGVSGAGSLAVTG-----GG 347
 QY 108 VVD-----TGGANKLAVNIDDLTLTGTGTAAAGNAPALLFQGGQAAANTTYA 158
 Db 348 LVDAANIIITGNTGGNGSVRVSGADSTVKRSRLNVLGXGSMVTVEAGGAVKSRDGYVA 407
 QY 159 LGNITLGGANAGLTASDPDLVLPITLGNIDGGIITDNTDAAINGT--IGNTPAAOI 216
 Db 408 ----TGGSTSAVTG-----DGSWAMTGTFFGVASGAT--440
 QY 217 SIGASTLSLGGAV--IKATTTKLTNAAPVLTLTNANAVLTGAVDNTT---GGDDVGVNML 271
 Db 441 --GNVTVSNGGALRATGVTGLDLAGAGTWTITGAGSKVTAYVDNGTVNSGSDVDFGQG- 497
 QY 272 NGALSQVTGN-----IGNTNSLATISVG-----AGTATL---GGA 303
 Db 498 SGSLSVYNGGSLDAYNLYVGNALGSSGAVLVSGVSHVSDGLVMVYGNAGNGSVETGGA 557
 QY 304 VIKATTKL-TNAASVLTIT-----NAVLTGAVDNTT---GGDNVGVV-----342
 Db 558 SLAAPILLATAGSTGVLSIGAGSQATARSAGAVEARAIFAGAGNGSIVFNHSETGYTL 617
 QY 343 --NLSGALSQVTGNIGTNSLATINIGAGVATLDGAVIKATTK-----LTDASVLI--393
 Db 618 SADISCA-GRVVAEAGVT-TLSCNNYSVGGTTISAGMLKGTAKSFSGSGGVVNAELVVDG 675
 QY 394 ---FTNPVVVTGAIDNTGNANKGVVLTGASVTDNIGNTAFLAELVSVGAGLQIOGGVV 450
 Db 676 GTTSLNAISGTSFERTGDCN---LLLTGNSTYS---GATA-----VSAGKLSVNGSL- 722
 QY 451 KANAILNTDNASVVTFTGDSV--TGSIGGTTELFATFVNIGAGITLRAGGSLAANNIDFGA 508
 Db 723 -----ASAVSVGSGATVGGTGTIG-----GLTVNSGGTLAPGN-----S 756
 QY 509 ASNLEFNGPAGKNYLNIGTIANNNATLINAAGT-----VIANDVSI 551

Db 757 IGTLTSTGNA-----TFASGSYAVEIDADGSSDRLAVTGTITIANDVSL 801
 RESULT 11
 US-09-167-568-61
 ; Sequence 61, Application US/09167568
 ; GENERAL INFORMATION:
 ; APPLICANT: LOOSMORE, Sheena M.
 ; APPLICANT: YANG, Yan-Ping
 ; APPLICANT: KLEIN, Michel H.
 ; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS
 ; NUMBER OF SEQUENCES: 91
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: 6th Floor, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/167,568
 ; FILING DATE: 07-OCT-1998
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEWART, Michael I.
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-828 MIS:jb
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 61:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1180 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-09-167-568-61

Query Match 10.7%; Score 322; DB 15; Length 1180;
 Best Local Similarity 25.2%; Pred. No. 1.3e-17;
 Matches 159; Conservative 83; Mismatches 251; Indels 138; Gaps 30;
 QY 48 AELAACTNIGPGAGAF-VAGSTLQYTGAFV-----TDADVSRLDLNPFAGLFSVT 100
 Db 543 AEIQIGNLSQREGLNLTSSDKINIKQITIKKGVNGENSSTKS-----QANLTIKT 596
 QY 101 GDISLGSVVDGANKLAVNIDDLTLTGTGTAAAGNAPALLF-----OGGQA 150
 Db 597 KELKLTQDLNISGNKAKIVAKDSSNLTGNSDSDGNTSAKTVTFNNVKDSKISADGHV 656
 QY 151 AANVTYALG---NITLGA--NAGLTI-ASDPDLVLPITLGNID---GGGIITDNTDA 201
 Db 657 TLNSKVKTLSDNDNTEGGSDNNTGLTITAKOVENVNNTTSHKTVNVSAANGGITTKGT 716
 QY 202 AINGTIGNTPAAQISIGASTLSLGGAVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTT 261
 Db 717 TINATAGN-----VEITAHTGSIQGI-----ESKP-----GSVTIVA 749
 QY 262 GGDVGVNLSGALSQVTGNIGTNSLA-----TISVGAGTATLGGAV-----IKA 307
 Db 750 GGDTLAVGNISGNAVTVTANSAGLTTLAGSTIKGTESITTSQSGNIGKISGKTVNVKA 809
 QY 308 TTTKLTNAASVLTTLTNAVLT-----TCGAVDNTTGGDNVGVNLSGALSQVTGNIGTNSLA 362
 Db 810 TNSLTQADSKIBATEGEANVTSKTSIIIGTISGGTVEVTATGLTQAGSTIITGESVT 869

[illegible]

```

RESULT 12
US-09-167-568A-61
: Sequence 61, Application US/09167568A
: GENERAL INFORMATION:
: APPLICANT: LOOSMORE, Sheena M.
: APPLICANT: YANG, Yan-Ping
: APPLICANT: KLEIN, Michel H.
: TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS
: TITLE OF INVENTION: INFLUENZAE HIGH MOLECULAR WEIGHT PROTEINS
: NUMBER OF SEQUENCES: 91
: CORRESPONDENCE ADDRESS:

```

```

; ADDRESS: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/167,568A
FILING DATE: 07-OCT-1998
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-828 MIS:jfb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 1180 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-167-568A-61

```

Query Match	10.7%	Score 322;	DB 15;
Best Local Similarity	25.2%	Pred. No. 1.3e-17;	Length 1180;
Matches 159; Conservative	83;	Mismatches 251;	Indels 138; Gaps 30;

[illegible][illegible]

RESULT 13
US-09-167-568-59
Sequence 59, Application US/09167568
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS
INFLUENZAE HIGH MOLECULAR WEIGHT PROTEINS
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/167,568
FILING DATE: 07-OCT-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-828 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:

LENGTH: 1188 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-167-568-59

Query Match 10.7%; Score 322; DB 15; Length 1188;
Best Local Similarity 25.2%; Pred. No. 1.3e-17;
Matches 159; Conservative 83; Mismatches 251; Indels 138; Gaps 30;

```
QY 48 AELAACTNIGCAGAF-VAGSTLQYTGATV-----TDADVSVRALDLNFAAGLSVTV 100
DB 551 AEIQGGNISKEGRLTSSDKINITKQITIKKGVNGENSSTKS-----QANLTIKT 604
QY 101 GDISGVVDGKANKLAVNDDGLTLTGTGTAAYGANPALLF-----QGGQA 150
DB 605 KELKTQDLNIGSPNKAIVAKDSSNLTIGNSDDSGNTSAKVTTFNNVKDSKISADGHV 664
QY 151 AANNTYTAG---NITLGA--NAGLTI-ASDPDVLGPITLAGNID---GGGIITDNTDA 201
DB 665 TLNSKVKTLSDNDNTEGSDNNTGLTITAKDVEYNNITSHKTVNVSAANGIITKTGT 724
QY 202 AINGTIGNTPAAQISIGASTLSLGGAVIKATTTKLTNAAPVLTNTNANAVLTGAVDNTT 261
DB 725 TINATAGN-----VEITAHTSGIOGGI-----ESKP-----GSVTIVA 757
QY 262 GDDVGVNLNICALSQVTGNIGNTNSLA-----TISVGAGTATLGGAV-----IKA 307
DB 758 GGDTLAVNIGSNVAVTAVNSGALTTLAGSTIKGTESITTSQSGNIGKISGKTVNVKA 817
QY 308 TTTKLTNAASVLTNLNAV-----TCAVDNTTGGDNGVNVNLSGALSQVTGNIGNTNSLA 362
DB 818 TNSLTQADSKLEATEGEANVTSKTSIIGGTISGGTVEVTATEGLTQAGSTITGESVT 877
QY 363 TI-----NIGAGVATLDGAVIKATTK--LTDDASVLIET---NPVVVTGAIDNTGNANK 412
DB 878 TSSQSGNIG--GMISGGKVEVSATKDLITKSGSEIKATAGEVNVTSATGTIDGTISGNT 934
QY 413 -GVVIFTGASTVDNIGNTAVLAESVAGLLQIOGG--VVKANALNTDNASVVTFTGD 469
DB 935 VNVNTANGDLTVED-----AAKIDATGAATLTATSGKLTTKASSITSANNQVNLAKD 989
QY 470 STVTSIGGTELFATVNIAGITLRAGGSLAANNIDFGA---ASNLEFNGPAGKYNLI 525
DB 990 GSIIGNINAANY--TLNITGALTTVKGSINANS---GTLVINAKDAELNGEASGHTVV 1044
QY 526 GTIANGNNATLINAGTVIANDVSIQTVQAINQNNKIFVINAKNADVDILDAQAISFK 585
DB 1045 -----NAT-NANGSGSVIAT-----TSSRYNITGD-LITINGLN-----ITSK 1080
QY 586 GAASRLFLANVSLQMIELSLKIIYVPLTV 616
DB 1081 NGINTVLLKGVKID-----VKYIQPGIASV 1105
```

RESULT 14

US-09-167-568A-59

Sequence 59, Application US/09167568A

GENERAL INFORMATION:

APPLICANT: LOOSMORE, Sheena M.

APPLICANT: YANG, Yan-Ping

APPLICANT: KLEIN, Michel H.

TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS

TITLE OF INVENTION: INFLUENZA HIGH MOLECULAR WEIGHT PROTEINS

NUMBER OF SEQUENCES: 91

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: 6th Floor, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/167,568A
FILING DATE: 07-OCT-1998
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: STEWART, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-828 MIS:jb

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 1188 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-167-568A-59

Query Match 10.7%; Score 322; DB 15; Length 1188;

Best Local Similarity 25.2%; Pred. No. 1.3e-17;

Matches 159; Conservative 83; Mismatches 251; Indels 138; Gaps 30;

```
QY 48 AELAACTNIGCAGAF-VAGSTLQYTGATV-----TDADVSVRALDLNFAAGLSVTV 100
DB 551 AEIQGGNISKEGRLTSSDKINITKQITIKKGVNGENSSTKS-----QANLTIKT 604
QY 101 GDISGVVDGKANKLAVNDDGLTLTGTGTAAYGANPALLF-----QGGQA 150
DB 605 KELKTQDLNIGSPNKAIVAKDSSNLTIGNSDDSGNTSAKVTTFNNVKDSKISADGHV 664
QY 151 AANNTYTAG---NITLGA--NAGLTI-ASDPDVLGPITLAGNID---GGGIITDNTDA 201
DB 665 TLNSKVKTLSDNDNTEGSDNNTGLTITAKDVEYNNITSHKTVNVSAANGIITKTGT 724
QY 202 AINGTIGNTPAAQISIGASTLSLGGAVIKATTTKLTNAAPVLTNTNANAVLTGAVDNTT 261
DB 725 TINATAGN-----VEITAHTSGIOGGI-----ESKP-----GSVTIVA 757
QY 262 GDDVGVNLNICALSQVTGNIGNTNSLA-----TISVGAGTATLGGAV-----IKA 307
DB 758 GGDTLAVNIGSNVAVTAVNSGALTTLAGSTIKGTESITTSQSGNIGKISGKTVNVKA 817
QY 308 TTTKLTNAASVLTNLNAV-----TCAVDNTTGGDNGVNVNLSGALSQVTGNIGNTNSLA 362
DB 818 TNSLTQADSKLEATEGEANVTSKTSIIGGTISGGTVEVTATEGLTQAGSTITGESVT 877
QY 363 TI-----NIGAGVATLDGAVIKATTK--LTDDASVLIET---NPVVVTGAIDNTGNANK 412
DB 878 TSSQSGNIG--GMISGGKVEVSATKDLITKSGSEIKATAGEVNVTSATGTIDGTISGNT 934
QY 413 -GVVIFTGASTVDNIGNTAVLAESVAGLLQIOGG--VVKANALNTDNASVVTFTGD 469
DB 935 VNVNTANGDLTVED-----AAKIDATGAATLTATSGKLTTKASSITSANNQVNLAKD 989
QY 470 STVTSIGGTELFATVNIAGITLRAGGSLAANNIDFGA---ASNLEFNGPAGKYNLI 525
DB 990 GSIIGNINAANY--TLNITGALTTVKGSINANS---GTLVINAKDAELNGEASGHTVV 1044
QY 526 GTIANGNNATLINAGTVIANDVSIQTVQAINQNNKIFVINAKNADVDILDAQAISFK 585
DB 1045 -----NAT-NANGSGSVIAT-----TSSRYNITGD-LITINGLN-----IISK 1080
QY 586 GAASRLFLANVSLQMIELSLKIIYVPLTV 616
DB 1081 NGINTVLLKGVKID-----VKYIQPGIASV 1105
```

RESULT 15

US-09-167-568-35
 ; Sequence 35, Application US/09167568
 ; GENERAL INFORMATION:
 ; APPLICANT: LOOSMORE, Sheena M.
 ; APPLICANT: YANG, Yan-Ping
 ; APPLICANT: KLEIN, Michel H.
 ; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS
 ; TITLE OF INVENTION: INFLUENZA HIGH MOLECULAR WEIGHT PROTEINS
 ; NUMBER OF SEQUENCES: 91
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: 6th Floor, 330 University Avenue
 ; CITY: Toronto,
 ; STATE: Ontario,
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/167,568
 ; FILING DATE: 07-OCT-1998
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEWART, Michael I.
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-828 MIS:jb
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 35:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 915 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-09-167-568-35

Query Match 10.5%; Score 316.5; DB 15; Length 915;
 Best Local Similarity 23.8%; Pred. No. 2.7e-17;
 Matches 153; Conservative 92; Mismatches 257; Indels 141; Gaps 26;

QY 42 TVTADGAELAGTNGPCAGAF-VAGSTLQYTGAFVTDADVSVRALDLNN-FAAGLFVS 99
 DB 229 TNGDNTETIQGGNISQEGNLTSSDKVNITERITI-KAGVNGDSDSNEATSANLTIK 287
 QY 100 TGDISLGSVVDGTGANKLAVNIDDLTLTLTGCTAAYGANPALLFOGGQAAANNYYTAL 159
 DB 288 TKELKLTNDLNLISGFNKAETAKDNSNLT-----GDNSDAGNTDAK- 329
 QY 160 GNITLGGANAGLTASDPDLGPTILAGNIDGGIITDNTDAAINGTIGNTPAAQISIG 219
 DB 330 -KVTFSNVKDSKISASDHNV---TLNSKVETSG-DTSDTEGCGNNNTGLTIKNTVN 383
 QY 220 ASTLSLGGAVIKATTKLTNAAPVLTITNANAVLTGAVDNTGGDDVGVNLNGLASQVT 279
 DB 384 NNITSHKTVNITASENVTTRAGTTINATGTSVEVTAKTGDKIGIESNGSNVNITASGDT 443
 QY 280 GNIGN-TNSLATISVGAG-TATLGGAVIKATTKLTNAASVLTITNAV-----LTCAVD 331
 DB 444 LNVSNITGNQNTVAAASGAVTTTKGTINAT-----TGNANITTKTGEINKEVKSASGNV 499
 QY 332 NTYGGDNVGVNLSGALSQVTGNIGNTNSLATINIGAGVATLDGAVIKATTKLTDDASV 391
 DB 500 ITASGNTLVNSITGNQNTVTVTANS-----AITTEGSTINAT---TGDANI 543
 QY 392 LIFTNPVVVTGAIDNTGNANKGVV--IFTGASTVTDNIGNTAVLAESVCGALIQO--- 446
 DB 544 -----TTQGNINGKVESGSSGVTLIATQTLAVNGISGDTV--TITADKGLTTQTSS 595
 QY 447 -----GGVVKANAINLTDNASVVTFTGD-----STVTGSIGG 478

Db 596 KINGTKSVTTSSQSGDISGTISGNTVSVSATGSLTTOAGSKTEAKTGEANVTSATGTIGG 655
 QY 479 TELFATVNI-----GAGITLRA-GGSLAANNIDFGAASNLEFNGPAG 519
 Db 656 TISGNTVNVNTANDNLTIKDGARIKATGAVTLTATGTLTETSSDITSSNGQTTLTA- 714
 QY 520 KYNLIGTIANGNNATLNINAAGTIVAN---DVSIGTVAQINIQNNKIFVINAKKADVD- 575
 Db 715 KDSSIAGSI-NAANVTLANTTGTLTVAGSKIEAASGTL-----VINAKQAQLDG 762
 QY 576 -----ILDAQAISFKGA--ASRLFLANVLSQMIELSLLKII 609
 Db 763 AASGDHTVVNATNANGSGSVIATTSRVNITGDLITINGLNII 805

Search completed: April 14, 2003, 16:14:42
 Job time : 165 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:09:29 ; Search time 82 Seconds
(without alignments)
1176.179 Million cell updates/sec

Title: US-09-800-065-2

Perfect score: 3022

Sequence: 1 MANISLKLQRAIKGLKTA.....MIELSLKLIYPVLLTVVVS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 704720 seqs, 155810517 residues

Total number of hits satisfying chosen parameters: 704720

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_New.*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	392	13.0	1862	6	US-10-282-122A-49757
2	373	12.3	979	6	US-10-282-122A-47372
3	366.5	12.1	1331	6	US-10-282-122A-47930
4	362	12.0	2457	6	US-10-282-122A-49854
5	353.5	11.7	827	6	US-10-282-122A-47312
6	350	11.6	4327	6	US-10-369-493-10178
7	342.5	11.3	3286	6	US-10-282-122A-49697
8	341.5	11.3	1407	6	US-10-282-122A-47709
9	331	11.0	1026	6	US-10-223-597-7
10	329.5	10.9	3705	6	US-10-282-122A-77944
11	322	10.7	1180	6	US-10-193-764-61
12	322	10.7	1188	6	US-10-193-764-59
13	319.5	10.6	3073	6	US-10-282-122A-49147
14	316.5	10.5	915	6	US-10-193-764-35
15	316.5	10.5	1222	6	US-10-193-764-37
16	316.5	10.5	1228	6	US-10-193-764-34
17	315.5	10.4	1129	6	US-10-282-122A-48048
18	308	10.2	2056	6	US-10-282-122A-50912
19	306.5	10.1	1074	6	US-10-282-122A-50616
20	301	10.0	1215	6	US-10-369-493-10073
21	298	9.9	3013	6	US-10-282-122A-78257
22	293.5	9.7	969	6	US-10-193-764-32
23	293.5	9.7	975	6	US-10-193-764-30
24	293	9.7	1465	6	US-10-369-493-13955
25	293	9.7	6310	6	US-10-282-122A-67793
26	289	9.6	1220	6	US-10-193-764-28

27	289	9.6	1226	6	US-10-193-764-26	Sequence 26, Appl
28	286	9.5	361	6	US-10-282-122A-47521	Sequence 47521, A
29	280	9.3	1036	6	US-10-193-764-69	Sequence 69, Appl
30	280	9.3	1477	6	US-10-193-764-67	Sequence 67, Appl
31	277	9.2	518	6	US-10-282-122A-50634	Sequence 50634, A
32	277	9.2	2044	6	US-10-282-122A-70017	Sequence 70017, A
33	276.5	9.1	1099	6	US-10-282-122A-48220	Sequence 48220, A
34	276	9.1	1407	6	US-10-312-273-31	Sequence 31, Appl
35	276	9.1	1407	6	US-10-282-122A-54678	Sequence 54678, A
36	273.5	9.1	1095	6	US-10-193-764-65	Sequence 65, Appl
37	273.5	9.1	1536	6	US-10-193-764-63	Sequence 63, Appl
38	273	9.0	1749	5	US-09-640-419C-28	Sequence 28, Appl
39	271	9.0	1246	6	US-10-282-122A-49773	Sequence 49773, A
40	269	8.9	1005	6	US-10-193-764-41	Sequence 41, Appl
41	269	8.9	1011	6	US-10-193-764-39	Sequence 39, Appl
42	264.5	8.8	1070	6	US-10-282-122A-78305	Sequence 78305, A
43	263	8.7	1649	6	US-10-369-493-18460	Sequence 18460, A
44	262.5	8.7	1333	6	US-10-282-122A-55546	Sequence 55546, A
45	262	8.7	992	6	US-10-193-764-57	Sequence 57, Appl

ALIGNMENTS

RESULT 1

US-10-282-122A-49757
; Sequence 49757, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49757
; LENGTH: 1862
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49757

```
Query Match      13.0%; Score 392; DB 6; Length 1862;
Best Local Similarity 28.0%; Pred. No. 5.5e-15;
Matches 161; Conservative 66; Mismatches 197; Indels 152; Gaps 26;

Qy 13 IOKGLKTAFTSTAAIMLTGSGVVGAAARTVTADGAELAACTNIGPGAGAFVAGSTLOYT 72
Db 1348 ITTGDLDSL-GASTSAV---GDVTLNAGRNA-----LNGTLVGGNGYISAGQDVTPG 1397
Qy 73 GAFTVTDADVSRALDLNFAAGLFSVTGDISLGSVDVTGGANKLAVNIDDLGLTLTLTGT 132
Db 1398 GTQAFTHAVAL-----GAGGDIALGSLQANSVQATGGDRAALNNVTSTTLTLTAN 1449
Qy 133 GTAAYGANPALLFOGQQAANNVTYALNITLGGANAGLTIASDPDVLGPTITLAGNIDGG 192
Db 1450 GNAG-----NGDAAITGTATAPGLVTLNAR-----DVL-----ISGALGGG 1486
Qy 193 GIIT--DNTDAAINCT---IGNTPAAOISIGASPLSLGGAVIKATTTKLTNAAAPVLT 247
Db 1487 TTVTLAQRNATVSGTQAVGDL-----VTAT 1514
Qy 248 NANAVLTGAVDNTTGGDDVGVNLNGAL--SQVTGNIGNTNSLATSISVG----- 294
Db 1515 NGSAALTGTATTT-----GALNVTSGLDLTQGGQIAVTNSL-TVQACTDVELTGTVAGQ 1567
Qy 295 -AGTATLGAIVKATTTKLTNAAVSLTLTNAVLTCGAVD-----NTTGGDNVGVNLSGALS 349
Db 1568 AAGSLTAGQDITGGGSAFAQAATLSAANNVALTGALGSGISVTGGNAGL-----GSAQ 1623
Qy 350 QVTGNIGNTNSLATINICAGVATLDGAV--IKATTTKLTDDASVLIFTPVVTGAIDNT 407
Db 1624 ATTGNLDVT---ANGNAGQDVTGGAVTAIGTTTLQARVDS-----VAGAINSG 1671
Qy 408 G-----NANKGVVIFPGASTVDNIGNTPAVLAESVSGAGLILQIOGVVYKANAINTLD 459
Db 1672 GTASITAAARNATAADVNSAGDLTITATSGNLSA-GNVTTOGNLINESAGQALSASQOTEG 1730
Qy 460 NASVVTFPGDSVTVGSGIGTFLFATVNIAGITLRAGSLAANNIDFGAASNLEFNGPAG 519
Db 1731 GNATLASSGDMTLGGI-----AAQNTG---TLTSGTLTAGSVAFQQQTINADG--- 1778
Qy 520 KYNILIGTIA-NGNNATLINAAAGTVIANDVSIGTV 554
Db 1779 -SINVSGSVAINGNLA-----ATATDSLMSGV 1805

RESULT 2
US-10-282-122A-47372
; Sequence 47372, Application us/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
```

```
Query Match      12.3%; Score 373; DB 6; Length 979;
Best Local Similarity 26.2%; Pred. No. 2.9e-14;
Matches 174; Conservative 85; Mismatches 278; Indels 128; Gaps 29;

Qy 16 GLKTALEFTTAAIMLTGSGVVGAAARTVTADGAELAACTNIGP--GAGAFVAGSTLOYTG 73
Db 341 GLNLLNNVTLGADLTVDGATGLALSGTIAGGALKASCTGTLTNGANSFTGGTNLTGGG 400
Qy 74 AFTV-TDADVSRALDLNFAAGLFSVTGDISLGSVDVTGGANKLAVNIDDLGLTLTLTGT 132
Db 401 GLVGTDTALGGGALNVSGAGGTGTSVGGTLLGN-----AINIDGTVLGL--N 448
Qy 133 GTAAYGANPALLFOGQQAANNVTYAL-----GNITLGGANAGLTIASDPDVLGPTIL 185
Db 449 GSAGLGSAGTAGNGLAQTFGITTLLGPNTFAGTTLSG---GGIVAGTGAALG--TG 503
Qy 186 AGNIDG-GGIITDNTDRAINGTIGNTPAAOISI--GASTLSLGGAV-----IKATTTKLT 238
Db 504 ALHVTGTGSLATNAGGTLLGNVNLDAAGATLGLNGANGLSISGTTIAGNGGLAQSGTGCT 563
Qy 239 NAAPVLITNANAVLTGAVDNTTGGD-DVGVNLNGALSQVTGNIGNTNSLATISVAGCT 297
Db 564 TLLGANTFSGGTALSGGGGLIAGNCAALGTGALNVSGTGGTTLATNVGTVLGNVNLGAGA 623
Qy 298 A-----TLGGAVIKATTTKLTNAAVSLTLTNAVLTCGAVDNTTGG-----DNVG 340
Db 624 TLVNGAVDLSLGAIGSGGQLAQTGAGTTTLLGTNTFTGTTLSGGGLVAGNCAALGTG 683
Qy 341 VVNLGALSQVTGNIGNTNSLATINIGAGVAT-LDGAIVIKATTTKLTDDASVLIFTPNV 399
Db 684 ALNVSGTGGTLATNVGTVLGNVNLGAGATLGLNGAA-----DLG 724
Qy 400 VTGAIDNTGN---ANKGVVIFGTAGTVTDN---IGNTAVLAE-----VSVGAGLILQIOG-- 447
Db 725 LSGAISGAGGLALGCTGATTLSSGANTTGGASLTGGGLIVEGGSALGIC-GALNVSGLG 783
Qy 448 -----GVVKANAINLTDNASVVTFTG--DSTVTSIGG-----TELFATV 485
Db 784 GSLRADLAAGAALANGINL--NGASLTILINGANDLGLSAICGTGGLVLSGTTGTTLSGAN 841
Qy 486 NIGAITLRRAGGSAAANNIDFGAASNLEFNPGACKNYNLICTIANG-----NNATLINAA 542
Db 842 TFGGTTLSAGTLLAGSNTALGSGA-LNVAGTGG---ALGASVNGTALGNA-VNLGAGA 895
Qy 543 TVI---ANDVSI-----GTVAQINTONNKIFVINAKNADVLDLDAQAIKFKGASRLF 592
Db 896 TLALTGANDUGLNGVIAGSGGLAIGSGITTLTSGANTFSGGTTLSGGGLLAGNAGALGTG 955
Qy 593 LANVS 597
Db 956 ALNVT 960
```

```
US-10-282-122A-47372
; ORGANISM: Burkholderia cepacia
; TYPE: PRT
; LENGTH: 979
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47372
```

RESULT 3

US-10-282-122A-47930
Sequence 47930, Application US/10282122A
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
PRIOR FILING DATE: 2003-02-20
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 47930
LENGTH: 1331
TYPE: PRT
ORGANISM: Burkholderia cepacia
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (101)..(101)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (236)..(236)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (254)..(254)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (256)..(256)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (268)..(268)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (282)..(282)
OTHER INFORMATION: X-any amino acid
FEATURE:

NAME/KEY: MISC_FEATURE
LOCATION: (295)..(295)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (303)..(303)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (310)..(310)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (312)..(312)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (324)..(324)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (461)..(461)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (468)..(468)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (488)..(488)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (498)..(498)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (528)..(528)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (577)..(577)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (942)..(942)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (944)..(944)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (962)..(962)
OTHER INFORMATION: X-any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (972)..(972)
OTHER INFORMATION: X-any amino acid
US-10-282-122A-47930

Query Match 12.1%; Score 366.5; DB 6; Length 1331;
Best Local Similarity 25.7%; Pred. No. 1.1e-13;
Matches 185; Conservative 65; Mismatches 235; Indels 235; Gaps 33;

QY 23 TTSTAIMLTGSG-----VLGAARTVADGAELAAGTNGICPGAGAFVAGSTLOYTGAF 75

Db 584 TLTGANTYTGTTINAGTLALGAGGSLASTGAMNLGAVGATLDLGATGGQTLGALSGV 643

QY 76 TTTDADSVRALDLNNFAAGLFSVTGDISLGSVVDTGGANKLAVNIDDDGLTLTLTGTT- 134

Db 644 AGTNINLGNALTLGCTASGTG-----GAIGGTG-----SLTLACTGTQ 683

```

Qy 135 -----AAY--CANPALLFOGQQAANNYYTALGNITLGG-----NAGLTIASDPDV--- 179
Db 684 TLNGASTYSGGTN---LNSGGVLCNNTALGTGTNNVGAATLDTNASTLVANAVNLGTG 740
Qy 180 -----LG---PTLGNIDGGIITDNTDAANGTIGNT-NPAAOISIGASTLSLGGAVIK 231
Db 741 AALTLGGSNALDLSGATSGAGSLVKNGAATTLTGANTYTGTTINAGTLAVGAGGSUAS 800
Qy 232 ATTKTLKLNAPVLITNANAVLT-GAVDNTTGGDDVGVNLNG-----ALSQVTGNI 282
Db 801 TGAMNLASAGATLDSAAATGAQTILGALSGVAGTN---NLGGNALTILGGTASCTFSGAI 856
Qy 283 GNTSLATISVAGTAYL-----GGAVIKATTKLTNAAASVLTNLNAV-LTGA----- 329
Db 857 GGTGSLTL--AGTGTQTLSGANTYTGTTIGGGSTLALGAGGSLASTGTNLAGAGATFD 914
Qy 330 VDNTTGGDNGVWNLGSLAL-SQVT----- 352
Db 915 LRASGAGTIGA--LTGAAGTNTVTSAAATGXRXAATAATRSAAASPAAXRSPRAVRRX 972
Qy 353 -----GNIGNTSLATINIGA-----GVATLDGAVI 378
Db 973 RARTIPAAARSAAARWRWAGGSLASTGAVTLAGTATLDSLGAATGAQTIGTLAGA-- 1030
Qy 379 KATTKLTDASVLFTNPVVVTGAIDNTGN--ANKGVVIFTGASTVTDNIGNTAVL-- 433
Db 1031 GGTSVNLGANAALTINATNGTFFGAIGGTGGTVAGTGTQTLTGANTY---GGTTINGG 1087
Qy 434 AEVSAGLLOIQGVV-----KANAINLTDNASVVFT 467
Db 1088 STLALGAGGSLASGAVTLAGAGATFDVSGATAQTGLTGAAGTNVNLGANAALTNGS 1147
Qy 468 GDSVTGSGIGT--ELFA-----TVNIGAGITLRAGGSLAANN----- 503
Db 1148 GNGTGAIGGTGGVTFAGTGTQTLTGANTYTGTTINGGSTLALGAGGSLASGSLNL 1207
Qy 504 -----IDFGAASNLE---FNGPAGKNVNLGTTANGNNAATLNAACTVANDVSGTV 554
Db 1208 GTGATFDMASAGAQITGALSGVGTNVNL-----GAN-TLTLNGSG-----NNTFGGTI 1256

RESULT 4
US-10-282-122A-49854
; Sequence 49854, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```

```

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See File wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 49854
; LENGTH: 2457
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49854

Query Match          12.08; Score 362; DB 6; Length 2457;
Best Local Similarity 27.6%; Pred. No. 5.3e-13;
Matches 195; Conservative 84; Mismatches 263; Indels 164; Gaps 42;

Qy 3 NISLKFQKATQKGLKTALETTSTAAM-LTSGVGLGAARTVTAD-GAELAAAGTNIGPGA 60
Db 50 NVTVOAGQ-IVNAGSITAVQNLIAAAVQTLFNGCTLAANGNVNASAGTTLTANTITAGR 108
Qy 61 GAFVAGSTLQVTGATVTDADVSRALDLNFAAGLSVTCGD-----ISLGSVVD-TGG-- 113
Db 109 QAAVSAATFDNCGSSASAD-OFTLSATNLVNH-GGSITQTGNGATSVNVSGLMDNTGTTI 166
Qy 114 --ANKLAVN---IDGLTLTLTGTT-----AAYGANPALLFQG----- 147
Db 167 QTNSTDLALGPATLNDNGKTIASSGSGTSLVKTGTLSNNGGTIATNGALMTDGGAVSNRG 226
Qy 148 ----GQAAANNYYTALGNITLG--GA-----NAGLTIASDPDLGPTITLAGNI 189
Db 227 GTLAGQSSATLRLVSLDNSAGGYIGAHASVIDGTGLNNAGGTIQAD-DALAVSAQSVTN 285
Qy 190 DGGIITDNTDA-AIN--GTIGTNT-----PAAQISGASTLSLGGAVIKATT 234
Db 286 DGGTIANGGTGATTVNAAGALTNTNGLGNGNVSVSGASIDNSGGTITTAAGATTVOQS 345
Qy 235 TKLTNAAPVLTLT-NANAVLTGAVDNTTGGDDVGLN-----LNGALSOVVTGNIGNT-NS 287
Db 346 STLGNRAGMIQCTGNVSASAGCAIDNTGGQIEVDGTNATMQLLAASLDNTNGRVANTNG 405
Qy 288 LATISVG-----AGTATLGG---AVIKATTTKLTAASVLT-----L 321
Db 406 ATTISAAGITNSNTGGVAGAGAIGGVDVTINATLTSNTNGAQLVGTGHDLTFLNIAQFANN 465
Qy 322 TNAVLTGAVDNTTGGDNGVVNLSS-----ALSOVVTGNIGNT-TNSLATNIG 367
Db 466 TNAVLTGAVDNTTGGDNGVVNLSS-----ALSOVVTGNIGNT-TNSLATNIG 367
Qy 368 AG-VATLDGAV-----IKATTTKLTDASVLIETN-PVVVTG--AIDNTG--NANKGVVI 416
Db 526 TGLANOGGAIGSDQNLSVTTNQLTGDRIIAGNDGAVTVNGDYTLDTGTQIQANHDLT- 584
Qy 417 FTGASTVTDNTGNTPAVIAEVSAGLQIQGVVYKANAINTLDNASVVFTFGDSTVTGSI 476
Db 585 FTTSNGFT-NOGTLGAVNALTVAANVDNQAG-ADLNSNTTVNAA-----GSISNAGRI 637
Qy 477 GG-----TELFATVN---IGAGITLRAG-----GSLANNIDFGAASNL-----BENGPA 519
Db 638 EGDVSTHSAALVNVNATIVGNTVTLNAGSIANTGAAAA--FAATAVNLVSPGDISNTGG 695
Qy 520 KYNLIGTIANGNNATNLINA----AGTVIANDVSIQTV-AQINTQ 560
Db 696 ANIESLGLDISIAADATRDANGLLANSRSVNDQS--TIEAQGNIE 739

RESULT 5
US-10-282-122A-47312
; Sequence 47312, Application US/10282122A

```

```

; GENERAL INFORMATION:
; APPLICANT: Wang, Liansu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47312
; LENGTH: 827
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47312

Query Match 11.7%; Score 353.5; DB 6; Length 827;
Best Local Similarity 26.3%; Pred. No. 3.4e-13;
Matches 186; Conservative 87; Mismatches 259; Indels 175; Gaps 36;

QY 24 TSTAATMLTGGG--VLGAART---VTADGAELAACTNIGPGAGAFVAGSTLQYTGAFTV 77
DB 5 SGAGGVQSGSGTTTLCGANTYGGGTTLSGGSLVSGSNTALGSGA-----LNVNGTGL 58
QY 78 TDADVSVRALDNNFAAGLFSVTGDISLGSVVDTGANKL----AVNIDGL-----TL 127
DB 59 S--ASVNGTTLGN-----AVT--LGAGATLGLNCANDLGLSGTISGSGLAQTGAGTT 107
QY 128 TLGTGCTAAYGANPALLFOGGAANNTY-----TALGNITLG-----G 166
DB 108 TLGTNTYDGGTT---LSGGGLVAGNGSALGTGALNVTGAGGSLGTSVGGTTLGNVNLG 164
QY 167 ANAGLIAASDPDVLGPITLGNIDGGIIT-----TNTDAALNGTIGNTPAAQISIGAS 221
DB 165 AGATLVGAND-----LGLGGISGSGLSVSGPSTTLTGNTYTGNTIG-----GGS 215
QY 222 TISL--GGAVIKATTTKLTNAAPVLTLTANAVLTGAVDNTTGGDDVGVNLNGLALSQVT 279
DB 216 TLAVGAGGSLSAGSALDLAGTGAALDISAATTPTQSMLSGVAGTWN---NLGNTLTLA 272
QY 280 GNIGNTNSLATIS-----VGAGTATL-----CGAVIKATTTKL-----TNAASVLT 320
DB 273 GT--GNGYAGTIGGTGGLTMAGTGTETLTGNTYTGATTINSGLTAIGAGGSLSSASPN 331

```

```

QY 321 LTNAVLT-----CAVDNTTGGDNVGV---VNLSGALSOVNTGNTNSLATINIGAGVAT 372
DB 332 LTGAGATFDVSGATTPTGTTLSGVAGSTVNLGGN--NLTLGGTGNGTYGTTAGAGGSLT 390
QY 373 LDGAVIKATT-----TKLFDASVLFTNPVVVVTGAIDNTGNANKGVIFTGASTVT 424
DB 391 LSGTGTELTGNTNTVGTGTTLSGGGTLIAGSGSALGTGALNTSGAGT-----LAASPTG 445
QY 425 DNIGNTAVLAEYVSGAGLLIOGGVVKANAIL-----TDNASVVTFTGDSTVT 473
DB 446 TTLGNA-----VNLGAGSTLVGG---ANDLGLGAISGGGLAVNGPSTTTLTGANTYT 497
QY 474 GS--TGGTELFATVNIAGITLRAGSL-----AANNIDFGAASNLE-----FNGPAGKNYN 523
DB 498 GSTTITGGG---STLAVGAGGSLSAGSAIDLSGTGATLDLSAATSPTSALSGGTGNTVN 554
QY 524 LIG---TIANGNNAT-----LINAACTVI-----ANDVSIGTVA---QIN 558
DB 555 LGGNTLTLLGGAGSGTYGGVIGGTGTLTSGTCTETLTGTNTYTGTATTINSGLTLAGNGS 614
QY 559 IONNKIFVINAKNADVDILDAQAISFKGAASRLFLANVLSLOMIELSL 605
DB 615 LSSSPVSLTAAGATLDLSGAASPOSTGAISGVAGSTVNLGNNTL 661

```

RESULT 6

```

US-10-369-493-10178
; Sequence 10178, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10178
; LENGTH: 4327
; TYPE: PRT
; ORGANISM: magnetite-containing magnetic coccus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(4327)
; OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-10178

```

Query Match 11.6%; Score 350; DB 6; Length 4327;

Best Local Similarity 25.6%; Pred. No. 6.5e-12;

Matches 169; Conservative 90; Mismatches 236; Indels 166; Gaps 31;

```

QY 16 GLK--TALTTTSTAATMLTGGSGVLGAARTV-----TADGAELAACTNIGPGAGAFV 64
DB 3061 GLKDGAAALAAANTNS--YITGSYVSVYDVTATMAQLSAMDTATTGTLTYGAGLKDSAAALV 3119
QY 65 AGSTLQYTGAFTVTDADVS-----VRLD-----LNNFAAGLFSVTGDISLGS---VVD 111
DB 3120 ANTNSVVTGAVTVTVDAAATTAQLGAIDQDTTGTVNYSLAGIKDVTNSITIDSGNVANA 3179
QY 112 GGANKLAVNIDGLTTLTGTGTAAYGANPALLFOGQQA---NNYITALGNITLGG 166
DB 3180 GGA---TITVNDGIANLIITDAGTVTGTNRVTVTDAASQAQLSQIDNYTTGALKYVTIKD 3236
QY 167 ANAGLIAASDPDVLGPITLGNIDGGIITDNTDAINGT--IGNTPAAQISIGASTLS 224
DB 3237 AVAALVA-----NTNSYVTSYAVSVTDAASMAQLSALDQD 3272

```

QY 225 LGGAVIKATTTKLTNAAPVLTILTNANAVLTGAVDNTTGGDDVGVULNNGALSOVVTGNTGN 284
Db 3273 TTGTL---TYTKLTDVAVNL-VTNTNSVTGSV-NVTVSIDIATISQLSSIDANTTGSVTY 3327
QY 285 T---NSLATISVGACT---ATLGGAVIKATT-TKLT-----NAAS 317
Db 3328 TOIGDAATLATNAGNYVKATIHVTVDAAVIAQLTTIDGNTTGSLSVYTAGGVKDSAN 3387
QY 318 VLTTLNAVLTGAVDNTTGGDNGVGVNLSGALSQVVTGNTG-----NTNSLAT 363
Db 3388 LVVNTNSVVTGAV-NVSVTDTVSIAQLSAVDEYTTGTLTYGAGVKDSVANLLVNTNSVYT 3446
QY 364 -----INIGAGVATLDCAVIKATT-----TKLTDASVLIFTNPVVVTGAI-----DNT 407
Db 3447 GSYAVSIITDVASMANL-SAIIDQFTGTLNLYTKLSDTVSALVANTNSYVTVGSVNVVTITDA 3505
QY 408 GNANKGVVIFTGASTVTDNIGNCTAVLAESVSGAGLQIQGGVWKANAIN--LTDNASVVT 465
Db 3506 SMAN-----MSAIDQNTTGTTLTYTKLSDTAALAANTNSYVTVGSVNVVTIDNATVAQ 3557
QY 466 FTG-DSTVTGSGIGTELPATVNIAGITPLRAGGSLAANNIDFGAASNLEFNGPAGKNY-- 522
Db 3558 LTTVDAAATGTI---KVASVYDSSGNI-----SSNFAYVDGLGVSYVN 3597
QY 523 ---NLIG-TIANGNATLINAGCTVIANDYSIGTVAQINTONKKIFVINAKNADVDILD 578
Db 3598 ANDNVMAVTAQADAV-----TIADDDVLTITDTSSTTSSITQGSTFALSALVATAGADATYLD 3651
QY 579 A 579
Db 3652 A 3652

RESULT 7

US-10-282-122A-49697
; Sequence 49697, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49697
; LENGTH: 3286
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49697

Query Match 11.3%; Score 342.5; DB 6; Length 3286;
Best Local Similarity 25.1%; Pred. No. 1.2e-11;
Matches 176; Conservative 82; Mismatches 250; Indels 193; Gaps 34;
QY 23 TTSTAAMLTGSGVLGAARTVTADGAELAGNITGCPGAGAFV-----ACSTLQYT 72
Db 797 TLSCAATNVSGASVDNSNGTIEGDTLALSSGNSLNRGRLTQYCSADOTLSAGTLDNT 856
QY 73 GAFVTVADVSVRALDLNNFAAGLFSVTGD-----ISLGVVVDTGANKLA 118
Db 857 GGFVATNA-----TNLTVSQSVTNDTSGIQHAGSGTFLASAGTFLSNTGG----- 902
QY 119 VNIDDLGLTLTGTTGTAAYGANPALLFQGOAA--ANNYTAGLNITLGGANAGLTASD 176
Db 903 -NVVTNGALDVGSTAVSNQGS-----LSAKGDATVRAOSLONHAGSVVAGG-NUGANLA-- 954
QY 177 PDVLGPI-----TLAG---NIDGGIITDNTDAAINGTI-GNT-----NPAA 214
Db 955 ---GALANQSGTSLGATTTVSGSSV-DNS---NCTIEGNTLALSSGNSLNRGRLT 1004
QY 215 QISIGASTLSLGGAVIKATTTKLTNAAPVLTILTNANAVLTG-AVDNTGG---DDVGLVN 270
Db 1005 QYGSADQTLTLAGGALDNTGTTATNA-----ANLTVSQSVTNDTSGIQHAGTGTLN 1056
QY 271 LN---GALSQVVTGNTGNTNSL-----ATISVGAGTATLGGAVIKATTTKLTNAASVL---T 320
Db 1057 VTPGALSVDVAGQIATNGALIARSASLDNSNGTVSAGSQAQVDANTSLTNRGGTLYKAG 1116
QY 321 LTNAVLTGAVDNTGG---DNVGVNLSGALSQVVTGNTGNTNSLATINIGAGVATLDGAV 377
Db 1117 LT-ATTQGAFTDQGSQVQTDGNSLVNAGGALSNTSGTI-----SVNGASGNAATATVSASS 1171
QY 378 IKATTTKLTDDASVLIFTNPVVVTGATDNTGNANKGVVIFTGASTVTDNIGNCTAVLAESV 437
Db 1172 IDNTSGKLTNSGGA--TTTATATGTTNSAGTMGGNGDVTILGAGTLLTNTASAKFVAARVA 1229
QY 438 -----VGAGLL-----QIQGGV-VKANAINLTDNASVVTET 467
Db 1230 SLNVNVRVNSGGTIYGGTALNLSQGSASVINDSGREGGQDYSVRVASLANSAGALAN 1289
QY 468 GDSTVTGSGIGT-ELFATVNIAGI-----TLRAG 496
Db 1290 RDITASGVWSGDGTMTAGRNGLAVTGDYTNASNSLHADGNMTVSATGTLTNAAGTLAAS 1349
QY 497 GSALAANNIDFGAASNLEFNGPAGKNYNICTIAN-----CNNATLN---INAGTVIAND 548
Db 1350 GALSAGANVVNAGADINS-ASTTVNATGTLTNAGRIEGDSVTTISATLANTGTTLGNT 1408
QY 549 VSICTVAQINIQNNKIFVINAKNADVDILDQAISFKGAAS 589
Db 1409 VQVNA--DVQNT-----GAAAIAAQAQSLNIYASNS 1438

RESULT 8

US-10-282-122A-47709
; Sequence 47709, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel

```

Qy 325 -VUTG-----VDNTGDNVGVVNLGSLASQVGTGIG-----NTNLSLATINI 366
      |||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||
Db 945 VALSGAGTVIANGADVLIDSFTVGGGIA-----AGATSIAGVGGSGVRNLGANATIAL 1000
Qy 367 -GAGVAT---LDGAVIKATTKLTDASVLIFT-----NPVVVTGADNTGNKGVVI 416
      |||      |||      |||      |||      |||      |||      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db 1001 SGTQVATTGNGADGLTSGAGARIATDAATVVVVTAGADARGVSVSGA--DSTLTANGTTIA 1059
Qy 417 FTGAST---VTDNIGHTAVLAEVSVCAGLQIQGGVVKANAINLTDNASVVVFTGDSVT 473
      |||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||
Db 1060 TTGANAHAIIMDG-GATALLSGAKVTAS-GNAADGIVAQNGGRIADTGSLSASAAGSGAT 1117
Qy 474 GSTGGTFLFATVYNIAGIITLRAGGSLAANNIDFGAASNLEFNGPAGKYNLIGTIANGNN 533
      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||
Db 1118 ANSGGV--LALT--GTALTGATAGVLTSDTLANGATSSVLIDGGSVTSATGPAPFAARGGT 1173
Qy 534 ATLNIINAAGVPIANDVSIGTVQAQINQNKIFVIAKNAKDVIDLDAQAISEFKGAASRLFL 593
      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||
Db 1174 ADIAVRNGTVVTAGN---GTLLNLANGSNVTSASAVNLAGDI-----VSDASSTGNVFL 1225
Qy 594 ANVSLQMIELSLLKIYPVLLTV 616
      |||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||
Db 1226 ANGT-----TLTGKIDPVALT 1242

RESULT 9
US-10-223-597-7
; Sequence 7, Application US/10223597
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: Nemellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDE
; FILE REFERENCE: 08106/002003
; CURRENT APPLICATION NUMBER: US/10/223,597
; CURRENT FILING DATE: 2002-08-19
; PRIOR APPLICATION NUMBER: US/09/379,931
; PRIOR FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/614,377
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: US 08/194,290
; PRIOR FILING DATE: 1994-02-09
; PRIOR APPLICATION NUMBER: US 07/895,367
; PRIOR FILING DATE: 1992-06-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Fast-Seq for Windows version 4.0
; SEQ ID NO 7
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-10-223-597-7

Query Match 11.0%; Score 331; DB 6; Length 1026;
Best Local Similarity 27.3%; Pred No. 1,1e-11;
Matches 181; Conservative 70; Mismatches 202; Indels 210; Gaps

Qy 23 TTSTAAILMTGSGVYLGAART-----VTADG-----AELAGTNIGPGA 60
      |||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||
Db 413 TTTTGAAVTTGGTAVTVAQTAGNAVNLTLTQADVTVTGNSSTAVTAVTQTAAAT-----A 467
      |||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||
Qy 61 GAFVAGSTLOYTGAFTVTDADVS-----VRALDLNNEAAGL-----FSVTG-- 101
      |||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||
Db 468 GATVAG---RVNGAATITLSAAASATAGKIATVTLGSGFATIDSSALTVVNLSGTGTS 524
      |||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||
Qy 102 -DISLGSVVDTTGGANKLAVNIDBGLTTLTGTGTAAYAGANPALLFGQGAAANNFTYALG 160
      |||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||
Db 525 LGIGRGALTATPANTLTLVNG--LTTTGAIT-----DSEAAADDGFT--- 566
      |||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||
Qy 161 NITLGGAGNAGLTIAS-----DPDVLGPITLAGNTDGGG--LIITDNTDAAING-TIGNTWPA 213
      |||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||      :||
Db 567 TINIAGSTASSTASLVAADATTL-----NISGDARVTTISHTAAALTGTTVTSNPGA 619

```

QY 214 ---AOSIGASTLSLGA-----VIKATTKLTNAAPVLTLTNANAVLTGAVDNTTGGD- 264
Db 620 TLGALATG--LVFTGGAGROSLILGATTKAIVMGAGDDTIVSSATL-GAGGSVNGGDG 676
QY 265 -DVGVLNGLALSQVGTGNIGTNSLATISVGAGTATLG-----GAVIKATTK 311
Db 677 TDVLVANVNGSFSADPAFGGFETLRV-----AGAAQSHNANGFTALQLGATAGATT-- 730
QY 312 LTNA-----SVLTJTNNAVLTGAVDNTTGGDNVGVNLSGALSQVGTGNIGTNSLATINI 366
Db 731 FTNAVNVGLTVLAAPTCTTTVTLANATGTSDFENLTLSAALAAGTVA-LAGVETVNI 789
QY 367 GA---GVATLDGAVIKATTKLTDDASVLFTNPPVVTGALDNTGNANKGVVFTGAST 422
Db 790 AATDNTTAHVDTLQLQTSK-----SIVVTG-----NAGLNL----- 823
QY 423 VTDNIGNTAVLAESVVGAGLLQIOGVVVKANAILTNDASVVVTFDSTVTSIGGTGTELF 482
Db 824 -TNTGNTA-----VTSFDSAVTGTAPATVFSANTTVGEV----- 858
QY 483 ATVNIGAGITLRAGSLAANNIDFGA-ASNLEFNGPAGKNYNLIGTIANGNNATL-NINA 540
Db 859 VTRIGGAGADSLTSATANDTIIGGAGADTLVYTGTT-----DTFTGGTGADIFDINA 911
QY 541 AGT-----VIANDVSTGTVAQINQNNKIFVINAKNADVDIILDAQAISFKGA-ASRLFLAN 595
Db 912 IGTSTAFVTITDAVGG-----DKLDLVGISTNGAIDAGAFGA 949
QY 596 VSL 598
Db 950 VTL 952

RESULT 10

US-10-282-122A-77944
; Sequence 77944, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77944
; LENGTH: 3705
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-77944

Query Match 10.9%; Score 329.5; DB 6; Length 3705;
Best Local Similarity 24.8%; Pred. No. 8.8e-11;

Matches 165; Conservative 88; Mismatches 233; Indels 179; Gaps 33;

QY 23 TTSTAAILTSG-----VLGAARTVTADGAELAAGTNIIGP---GAGAFVAGST 68
Db 1144 TTVTAGELLGANGAFGQTSLLDIASGASANINGVSQVGVAVTVGTVTLGSGGVLTSG 1203
QY 69 LOYTGAFVTVDADVSRALDNLNFAAGLFSVTGDISLGSVVDTGANKLAV----- 119
Db 1204 LTNGGILDLTGG-----ALNLTAGGASTVAGGLTGAGTLNNGGN-LSVSAANSGLSG 1255
QY 120 --NIDDDGLTTLTGTTAAYGANPALLFQGGQAAANNYTTALGNITLGGANAGLTIASDP 177
Db 1256 QTHIADVASVTLTDTGTLGTS-----VEVLGTNLNNGANAMTNVLSG 1299
QY 178 D-----VLGPITLAGNIDGGIITDNTDAAINGTIGNTPAAQISIGASTLSGGAVIKAT 233
Db 1300 DGTINTNAAVTLSGNNSFSGAHQIGTDCGL--TVGQ---ASNLGASSATVNLG---TLT 1350
QY 234 TTKLTNAPVLTLTNANAVLTGAVDNTTGGDDVGVNLNGLALSQV-----TGN 281
Db 1351 SHLILNGVSESIANVLSGVAGSTVDIIGGADTALTANNSGFLGQYALAGNSKLTVASTNN 1410
QY 282 IGNTNSLA-----TISVGAGTATLGGAVIKATTTKLTNAAASVLTLT-----NAVLTGA 329
Db 1411 LGASSVALACAGDTLSLSGFNGTFGNSVGTSGVLQVTDRAEV-TLTSSNCVSNVAITDI 1469
QY 330 VDNTTGGDNVGVNLSGALSQVGTG---NIGNTNLSLATINIG-----AGVATL----- 373
Db 1470 ADATLDDLDIALFN-----HVLTGNGLLNVAKNADASTAFDFGTVGGAFSGIVNLTNTTF 1524
QY 374 ---DGAVIKATTKLTDDASVLFTNPPVVTCAIDNTGNA---NKGVVIFTGAS----- 421
Db 1525 ALSADNAAALARATLKLSDDS-----VTTVGATDRTLHGLDNGTGLTFIDGSPQSQ 1576
QY 422 ---TVDNIGNTAVLAESVVGAG-----LLQTQGGVW---KANAILTD 459
Db 1577 ANGVTVTDLALNSGTIS--ITGAGNWNENHPVTPPNVSLLEQDRGDLILLELINAANTVG 1634
QY 460 NASVVTFTGDSVTGSGIGTGFELFATVNVIGAGITLRAGGSLAANNI-DFGAASNLEFNPA 518
Db 1635 NANNLDDLVDGTAITS--GTQ-----GVESAIOQGGSTVANAHHNYGLTSS---NGNG 1682
QY 519 GK---NYNL--IGTIANGNNATLINAACTVIANDV-----SIGTVAQINONTNNKIFV 566
Db 1683 GSGLYVNTLSALELLANGANALLATESG-LTANRVNLNAELFCVGGGLV-VDAQNGALT 1740
QY 567 INAKN 571
Db 1741 ANGN 1745

RESULT 11

US-10-193-764-61
; Sequence 61, Application US/10193764
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Kleid, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764


```

; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 1180
; TYPE: prt
; ORGANISM: Haemophilus influenzae
US-1937-764-61

```

```

Query Match      10.7%  Score 322;  DB 6;  Length 1180;
Best Local Similarity 25.2%;  Pred. No. 4.5e-11;
Matches 159;  Conservative 83;  Mismatches 251;  Indels 138;  Gaps 30;

QY  48  AELAAGTNIGPGACAF-VAGSTLOYTGAFIV-----TDADVSVRALDNNFAAGLFSVT 100
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  543  AEIOIGGNISQKENLRISSDKINITKOITIKGVNGENSDDSYKS-----QANLTIKT 596
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY  101  GDISLGSVDPDGGANKLAVNIDDDLTLTLCTGCTGAAYGANPALLF-----QGQQA 150
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  597  KELKLTODLNIISGFNKAKIVAKDSNLTIGNSDDSGNTSAKTVTFNNVKDKSISADGHKV 656
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY  151  AANNTYFALG--NITLGGCA--NAGLTI-ASDPDVLGPITLAGNID---GGGIITDNTDA 201
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  657  TLNKKVKTLSNDNNTGEGSDNNTGLITAKDVEVNNNITSHKTVFNNVSAANGGITKTGT 716
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY  202  AINGTIGNTPAAQISIGASTLSLGGAVIKATTTKLTAAPVLTLTNANAVLTGAVDNTT 261
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  717  TINATAG-----VEITAHTGSIQGGI-----ESKP-----GSVTIVA 749
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY  262  GGDDVGVLNGLALSOVYTGNGINTNSLA-----TISVGAGIATYLGAV-----TKA 307
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  750  GGDTLAVGNISGNVAVTVANSALATTLAGSTIKCTESITTSQSGNIGCKIGSKGVTVYKA 809
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY  308  TTKTLTAASVLTLTNAVL-----TGAVDNTTGGDNVGVNLSGALSQVGTGNGINTNSLA 362
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  810  TNSLTQADSKIEATEGEANVTSKTSIIIGTISGGTVEVTEGLTTQAGSTITGESVT 869
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY  363  TI-----NITGAGVATLDGAVIKATTK--LTDASVLIFT-----NPVVVTGAIDNTGNANK 412
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  870  TSSQSGNIGT---GMISGGKVESATKDLITKGSSEIKATAGEVNVTSATGTDIGTISGNT 926
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY  413  -GVVIFTGASVTDNIGNTAVTLAEVSVCGALLQIOGG--VVKANAINLTDNASVVFTGD 469
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  927  VNVVTANTGDILVED-----AAKIDATGAATLTATSGKLTYTKASSIITSANNQVLSAKD 981
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY  470  STVTSGISGGTELFATVNTIGAGTITRAGGSLAANNIDFGA---ASNLEFNGPAGKYNLI 525
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  982  GSIQGNINAANY--TLNNTTGALTTVKGGSSINANS---GTLVINAKDAELNGEASGNHTVV 1036
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY  536  GTIANGNATLINAAQGVPIANDVSVIGVVAQIINONNKFIYNAKNADVDILDAQAISFK 585
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  1037  -----NAT-NANGSGSVIAT-----TSSRVNITGD-LITINGLN-----IISK 1072
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY  586  GAASRLFLFANYSLQMIELSLKIIYPPVLT 616
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB  1073  NGINTVLLKGVKID-----VKYIOPGTASV 1097
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

```

```

RESULT 12
US-10-193-764-59
: Sequence 59, Application US/10193764
: GENERAL INFORMATION:
: APPLICANT: Loosmore, Sheena M.
: APPLICANT: Yang, Yan-Ping
: APPLICANT: Klein, Michel H.
: TITLE OF INVENTION: PROTECTIVE RECOMBINANT
: TITLE OF INVENTION: MOLECULAR WEIGHT PROTEIN
: FILE REFERENCE: 1038-1239MIS
: CURRENT APPLICATION NUMBER: US/10/193,764
: CURRENT FILING DATE: 2002-07-12
: PRIOR APPLICATION NUMBER: 09/167,568

```

```

; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 1188
; TYPE: prt
; ORGANISM: Haemophilus influenzae
US-10-193-764-59

```

Query Match	10.78;	Score 322;	DB 6;	Length 1188;
Best Local Similarity	25.28;	Pred. No. 4.6e-11;		
Matches 159;	Conservative	83;	Mismatches 251;	Indels 138; Gaps 30;
Qy	48	AELAAGTNIGPGAGF-VAGSTFLOYTGAFIV-----TDADVSVRALDNLNFAAGLFSVT	100	
Db	551	AEIQIGGNISQEGNLIPISSDKINIKQITIKKGVNGENSDDSYKS-----QANLTIKT	604	
Qy	101	GDISIGSVVDYGGANKIAVNIDDGILTLTGCTGAAYGANPALIF-----QGQQA	150	
Db	605	KELKTODNLNIGFNKAKIVAKDSNLTIGNSDDSGNITSKTVTFNNVKDSKISADGHKV	664	
Qy	151	AANNTYATLG--NTILGGA--NAGLTI-ASDPDLGPITLAGNID--GGIITLNDTDA	201	
Db	665	TLNSKVYKLSNDNNTGEGSNNTGLITAKDVEVNNNITSHKIVNVSAAANGIITKTGT	724	
Qy	202	AINTGIGNTPAAQISIGASTLSLGGAVIKATTKLITNAAPVLTLTANAVLTVGAVDNTT	261	
Db	725	TINATAGN-----VEITAHGTSGIOGGI-----ESKP-----GSVTVIA	757	
Qy	262	GGDDVGVNLNGLALSQVTCNTGNNTNSLA-----TISVGAGATLGGAV-----IKA	307	
Db	758	GGDTLAVGNISGNNAVTVTANSGLATTLAGSTIKGPESITTSQSGNICGKISGKTVMVYKA	817	
Qy	308	TTTKLTNAASVLTLTNAVLI-----TGAVDNTTGGDNVGNVNLGSGALSQVTCNIGNTNSLA	362	
Db	818	TNSLTQADSKIEATEGEANVTSKTSIIGGITSGTVEVTEGLTTQAGSTIIGTESVT	877	
Qy	363	TI-----NTGAGVATLDGAVIKATTK--LTDASVLIFT-----NPVVVTGAIDNTGNANK	412	
Db	878	TSSQSGNIG--GMSIGGKVEVFATKDLITKSGSIIKATAGEVNVTSATGIDTIGSNT	934	
Qy	413	-GVVIFTCASTVTDNIGNTAVTAVLAEVSGAGLLQIOGG--VVKANAINLTDNASVVTFGD	469	
Db	935	VNVVTANTGDLTVED-----AAKIDATGAATLTATSGKLTTKASSIITSANNQVNLSAKD	989	
Qy	470	STVTSGISGGTELFATVNTGAGITLIRAGGSLAANNIDFGA----ASNLEFNGPAGKNYMLI	525	
Db	990	GSIGGNINAANY--TLNNTTGALLTVTKGSSINANS---GTLVINAKDAELNGEASGNHTVV	1044	
Qy	526	GPIANGNATNLINAAGTVPIANDVSIGTVVAQIINNKNKIFVNAKNADVDILDQAIAISFK	585	
Db	1045	-----NAT-NANGSGSVIAT-----TSSRVNITGD-LTINGLN-----LISK	1080	
Qy	586	GAASRLFLANYSLQWIELSLKLIIPVLLTV	616	
Db	1081	NGINTVLLKGVKID-----VKYIOPGTASV	1105	

RESULT 13
US-10-282-122A-49147
; Sequence 49147, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudlo, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Onlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.

```
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 49147
; LENGTH: 3073
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49147
```

```
Query Match 10.6%; Score 319.5; DB 6; Length 3073;
Best Local Similarity 25.0%; Pred. No. 2.7e-10;
Matches 175; Conservative 86; Mismatches 251; Indels 189; Gaps 35;

Qy 23 TTSTAAT---MLTG-SGVLGAAATVTDGA-----ELAAG-----TNIGPGAGAFVAGST-- 68
Db 1031 TNGTDAVNVLGSVTSITGGGAANPDGTIKKPSFAIGGTYTDVGSAINAAVSGGTAN 1090
Qy 69 -LQVTGAFVTDADVSRALDNLNFAAGLFSVTGDISLGSVVDTGGANKLAVNIDD---- 123
Db 1091 GVQY-----DTSAR-----TKVTLGGTGATTAVILSNVAN-GVANNDRAVNTQLQAM 1136
Qy 124 GLTLTLTGTAAGA-----NPALLFOGGQAAANNYITLGNITLGGANAGLTIASD 176
Db 1137 CATICTSGVVTNSFVAYDDTTQGIKSLKGTGTTITNVKAGALSASLDVAVNGSOLYQTN 1196
Qy 177 PDVLGPITLAGNI-----DGGGIITDNTDAINGTIGNTNPAQAQISGASTLSL 225
Db 1197 ANV---ANVAGNVANVTNVNNTNGGKIYFHA-----NSTLADSS-----ATGNSVAI 1244
Qy 226 GGAVIKATTTKLTNAAPVLTLTNANAVLTGAVDN-----TTGGDDVGVLNLCAL 275
Db 1245 GGAA-SATAANSVALGANSVAGRANAVSGVGERQIIVANATNTGDVAVNLQOLQAMG 1303
Qy 276 SOVNTGNTGNTS-LATISVAGTATLGA-VIKATTTKLTNAASVLTITNAV-----L 326
Db 1304 ANVNSSGVVTSFVAYDDTSKGVTLGSGGFTKAVT--LTNVANGVANADAVNMAQLKAM 1361
Qy 327 TGAVDNTTGGDNVGNVLSGALSQV--GNIGNTNSLATINIGAVATLDG----AVIKATT 382
Db 1362 GGTIDSSGNVTNAFVAYDDTSKGVTLGSGAGSTKAVALTNVANGVANADAVNMAQLKA-- 1419
Qy 383 TKLDDASVLFTNPVVVVTGAIDNTGNANKGVVIF--TGASTVT-DNIGNTAVLAESVVG 439
Db 1420 -----MGFTIDSSGNVTNAFVAYDDTSKGVTLGSGSTKAVA----- 1457
Qy 440 AGLLIQGGVVVKANAINLT-----DNASVVT-----FTGDSVTT-GSTGGTTELF 482
Db 1458 --LTNVANGVANADAVNMAQLKAMGGTIDSSGNVTNAFVAYDDTSKGVTLGSGAGSTKAV 1515
```

```
Qy 483 ATVNIGAG-----ITLRAGGSAAAN 503
Db 1516 ALTNVANGVANADAVNMAQLKAMGGTIDSSGNVTNAFVAYDDTTLGKISLKGTTGTTTN 1575
Qy 504 IDFGAASNLFEFNGPAGKNYNLIGTIANGNNATLNT-NAAGTV--IANDVSIGTVAAQINIQ 560
Db 1576 VKAGALSSTSLDAVNGS--QLYQTNANVANVAGNVANVTNVN-----NIT 1626
Qy 561 NN---KIFVINAKNADVDILDAQAISFKGAASRLFLANVSL 598
Db 1627 NGGGIKYFHANSTLADSSATGANSVATGGAASATAANSVAL 1667

RESULT 14
US-10-193-764-35
; Sequence 35, Application US/10193764
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 35
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-35
```

```
Query Match 10.5%; Score 316.5; DB 6; Length 915;
Best Local Similarity 23.8%; Pred. No. 6.6e-11;
Matches 153; Conservative 92; Mismatches 257; Indels 141; Gaps 26;

Qy 42 TVTADGAELAAAGTNI GPAGAF-VAGSTLQVTGAFVTDADVSRALDNLN-FAAGLFSV 99
Db 229 TNGDNTETIQIGGNISKEGNLTISSKVNITERITI-KAGVNGDSDSNRATSANLTIK 287
Qy 100 TGDISLGSVVDTGANKLAVNIDDLTLTLTGCTAAYGANPALLFOGGQAAANNYITL 159
Db 288 TKELKLTNDLNLISGFNKAEITAKDNSNLT-----GDNSDAGNTDAK-- 329
Qy 160 GNITLGGANAGLTIASDPDVLGPITLAGNIDGGIITDNTDAAINGTIGNTNPAQAQISG 219
Db 330 -KVTFSNVKDSKTSASDHNV---TLNSKVETSG-DTSDSTEDGNNNTGLTITAKNVTN 383
Qy 220 ASTLSLGGAVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGDDVGVLNLCALSOVT 279
Db 384 NNITSHKTVNLITASENVTTKAGITINATTSVEVATKDGKIGIESNSGNVNTASGDT 443
Qy 280 GNIGN-TNSLATISVGAG-TATLGGAVIKATTTKLTNAASVLTITNAV-----LTGAVD 331
Db 444 LNVSNITGNQVTVAAASGAVTTTKGSTINAT---TCNANITTKTGEINGEVKSASGNV 499
Qy 332 NTTGGDNVGVNLSGALSQV--GNIGNTNSLATINIGAVATLDGAVIKATTTKLTDDASV 391
Db 500 ITASGNTLNVSNITGNQVTVTANS-----AITTEGSTINAT---TGANI 543
Qy 392 LIETNPVVVVTGAIDNTGNANKGVV--IFTGASTVTDNIGNTAVLAESVVGAGLIQI-- 446
Db 544 -----TTQGTNGINGKVESSSGVTLTATQTLAVGNISGDTV--TITADKGLTGTSS 595
Qy 447 -----GGVVVKANAINLTDNASVVTFTG-----STVTGSGIGG 478
Db 596 KINGTKSVTTSSSGDTSIGTISGNTSVSATGSLTTOAGSKIEAKTGEANVTSATGTIGG 655
Qy 479 TELFATVNI-----GAGITLRA--GSLAANNIDFGAASNLFEFNGPAG 519
```

Db 656 TISGNTVNTANTDNLTIKDARIKATGAVTLTATGTLTETSSDITSSNGQTLTA- 714
QY 520 KYNLIGTIANGNNATLNINAAGTVIAN---DVSIGTVAQINQNNKIFVINAKNADVD- 575
Db 715 KQSSIAAGSI-NAANTVTLNTGTLTAVGSKIEAASGTL-----VINAKDAQLDG 762
QY 576 -----ILDAQAISFKGA--ASRLFLANVSLQMIELSLLKII 609
Db 763 AASGDHTVVNATNANGSGSVIATTSRNVITGDLITINGLNII 805

RESULT 15
US-10-193-764-37
; Sequence 37, Application US/10193764
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 1222
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-37

Query Match 10.5%; Score 316.5; DB 6; Length 1222;
Best Local Similarity 23.8%; Pred. No. 1e-10;
Matches 153; Conservative 92; Mismatches 257; Indels 141; Gaps 26;

QY 42 TVTADGAELAAAGTNGIPGAGAF-VAGSTLOYTGAFVTADVSVRALDLNN-FAAGLFSV 99
Db 536 TNGDNTETIQIGNISQKEGNLTISSDKVNITERITII-KAGVNGDSDSNEATSANLTIK 594
QY 100 TGDISLGSVVDTGGANKLAVNIDDLGLTLTGTGTAAAGANPALLFQGGQAAANNYYTAL 159
Db 595 TREKLTNDLINSFGNKAETAKDNSNLTI-----GNSDAGNTDAK- 636
QY 160 GNITLGGANAGLTIASDDPVLGPITLAGNIDGGIITDNTDRAINGTGTNPAAQISIG 219
Db 637 -KVTFSNVXKSKISASDHNV---TLNSKVETSG-DTSDTEGNNNTGLTITAKNVTYN 690
QY 220 ASTLSLGGAVIKATTTKLTNAAPVLTNANAVLTGAVDNTTGGDDVGVNLNLGALSQVT 279
Db 691 NNITSHKTVNITASENVTTKAGTTINATGVSVEVTAKTGDIKGGIESNGNVNITASGDT 750
QY 280 GNIGN-TNSLATISVGAG-TATLGGAVIKATTTKLTNAASVLTNNAV-----LTGAYD 331
Db 751 LNVSNITGQNVTVAAASGAVTTTKGSTINAT-----TGNANITTKTGEINGEYKVSASGNVN 806
QY 332 NTGCGDNVGVNLSGALSQVGTGNIGNTNSLATINIGAGVATLDGAVIKATTTKLTDDASV 391
Db 807 ITASGNTLVNSNITGQNVTVTANS-----AITTEGSTINAT---TGDANI 850
QY 392 LIFTNPVVVTGADNTGNANKGVV--IFTGASTVTDNIGNTAVLAEVSVGAGLLOIQ--- 446
Db 851 -----TTQTGNGKNGVSSSVTLIATGQTLAVNIGSDTV--TITADKGLTTQTSS 902
QY 447 -----GGVVKANAINLTDNASVWFTTGD-----STVTGSIGG 478
Db 903 KINGTKSVITSSQSGDISGTISGNTVSVSATGSLTTQAGSKIEAKTGEANVTISATGTIGG 962
QY 479 TELFATVNI-----GAGITLRA--GSLAANNIDFGAASNLEFNGPAG 519
Db 963 TISGNTVNTANTDNLTIKDARIKATGAVTLTATGTLTETSSDITSSNGQTLTA- 1021

QY 520 KYNLIGTIANGNNATLNINAAGTVIAN---DVSIGTVAQINQNNKIFVINAKNADVD- 575
Db 1022 KQSSIAAGSI-NAANTVTLNTGTLTAVGSKIEAASGTL-----VINAKDAQLDG 1069
QY 576 -----ILDAQAISFKGA--ASRLFLANVSLQMIELSLLKII 609
Db 1070 AASGDHTVVNATNANGSGSVIATTSRNVITGDLITINGLNII 1112
Search completed: April 14, 2003, 16:16:10
Job time : 96 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:07:44 ; Search time 21 Seconds
(without alignments)
2833.677 Million cell updates/sec

Title: US-09-800-065-2
Perfect score: 3022
Sequence: 1 MANISLKLQRAIQKGLKTA.....MIELSLKLIYVLLTVVVS 619

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1510	50.0	2021	2 A97859	190-KDa cell surfa
2	1494.5	49.5	2249	2 A41477	190K surface antig
3	1103	36.5	1029	2 T30852	outer membrane pro
4	452.5	15.0	2340	2 B71704	cell surface antig
5	372.5	12.3	1651	2 JC1340	outer membrane pro
6	333.5	11.0	1026	2 A48995	paracrystalline su
7	333.5	11.0	1073	2 C87374	S-layer protein Rs
8	333	11.0	1643	2 D71630	outer membrane pro
9	331.5	11.0	1655	2 E97835	hypothetical prote
10	329.5	10.9	3705	2 AD0123	probable autotrans
11	329	10.9	2554	2 AB3528	extracellular seri
12	323.5	10.7	1052	2 AF2959	conserved hypothet
13	323.5	10.7	1341	2 H98323	hypothetical prote
14	310.5	10.3	652	2 E97857	cell surface antig
15	308.5	10.2	256	2 G97778	cell surface antig
16	302	10.0	1145	2 G87284	hypothetical prote
17	299.5	9.9	1645	2 JN0896	crystalline surfac
18	298	9.9	3013	2 AB0480	probable invasiv y
19	295.5	9.8	1369	2 T17504	hypothetical prote
20	293.5	9.7	1910	2 AF0394	probable adhesin h
21	286	9.5	820	2 T17519	cell surface antig
22	286	9.5	1477	2 B43855	high-molecular-wei
23	282.5	9.3	1335	2 T17508	glycoprotein Vp260
24	281.5	9.3	1109	2 A56143	surface-array prot
25	278	9.2	591	2 AC3528	extracellular seri
26	277	9.2	5188	2 B85547	probable RTX fami
27	277	9.2	5291	2 F90696	hypothetical prote
28	276.5	9.1	1300	2 S07575	outer membrane pro
29	276	9.1	1407	2 B72078	polymorphic outer

30	273.5	9.1	1536	2 A43855	high-molecular-wei
31	273	9.0	1749	2 S75138	hypothetical prote
32	270.5	9.0	1176	2 T18042	ice nucleation pro
33	270	8.9	962	2 AG2444	hypothetical prote
34	270	8.9	1158	2 AF1852	hypothetical prote
35	269.5	8.9	1280	2 AB1981	hypothetical prote
36	269	8.9	1461	2 E90696	hypothetical prote
37	269	8.9	1461	2 A85547	hypothetical prote
38	267.5	8.9	5627	2 C83339	hypothetical prote
39	266	8.8	918	2 AB2445	hypothetical prote
40	266	8.8	1428	2 AC2224	hypothetical prote
41	264.5	8.8	1070	2 A10484	probable hemagglut
42	263.5	8.7	3535	2 E83641	probable autotrans
43	263	8.7	1649	2 C86822	hypothetical prote
44	262.5	8.7	1268	2 B99789	hemagglutinin/hemo
45	262.5	8.7	1270	2 E85649	hypothetical prote

ALIGNMENTS

RESULT 1

A97859

190-KDa cell surface antigen [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: A97859
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.,
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: A97859
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-2021 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03811.1; PID:g15620410; GSPDB:GN00173
C:Genetics:
A:Gene: rompa

Query Match 50.0%; Score 1510; DB 2; Length 2021;
Best Local Similarity 37.3%; Pred. No. 6.4e-67;
Matches 402; Conservative 53; Mismatches 127; Indels 496; Gaps 18;

QY	1	MANISLKLQKALQGLKTALETTSTAALMTGSGVLG-AARTVTADGAELAAGTNI	GP 59
DB	1	MANISPLKFKALQGLKAALETTSTAALMTSSGALGIAVSGVIATNNNAFSDN	VGN 60
QY	60	---AGAFVAGST-----LQYTGAFVTVDADVSRALDINFA----AGLFSVTG	101
DB	61	WNEITAGVANGTTPARGPQNNAFTYGGDYTIT-ADVADHIITAINVADTPIGL	NIAQ 118
QY	102	DISLGSVDTGGANKLAVNIDDLGLTLTGTTGTAAYGANPALLFQGGQAAANTY	ALGN 161
DB	119	NTVGSIV-TGG-NLLPVTITAGKSLTLNGNADA--AN-----HGFGAPADN	-YTLGN 168
QY	162	ITLGGANAGITIASDPLGPIITLACNIDGGIITDNTDAIINGTICNTNPAQIS	IGAS 221
DB	169	IALGGANAALIQSAAP--AKITLACNIDGGIITDNTDAIINGTICNTNPAQIS	IGAS 226
QY	222	TLSLGGAVIKATTTKLTNAAPVLTJNANAVLTGAVDNTGGDDVGLNINLGALS	QVGTN 281
DB	227	IATLEGAAIKATTTKLTNAASVLTJNANAVLTGAVDNTGGDDVGLNINLGALS	QVGTN 286
QY	282	IGNTNSLATISVGAGTATLG-----	301
DB	287	IGNTNSLATISVGAGTATLG-----	301
QY	302	-----	301
DB	347	VFTGSDVTGNIGNTNSLATISVGAGKATLGGAIIKATTTKLTNDASAVTFTN	PVVVTG 406
QY	302	-----	301

Db 407 AIDNTGNANNGIVFTGDSVTGNTGNNTNALATISVGAGKATLGGAIKATTTKLTNDAS 466
Qy 302 -----
Db 467 AVTFTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGNNTNALATISVGAGKATLGGAI 526
Qy 302 -----
Db 527 KATTTKLTNDASAVTFTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGNNTNALATISV 586
Qy 302 -----
Db 587 GAGKATLGGAIKATTTKLTNDASAVTFTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGN 646
Qy 302 -----GAVIKATTTKLTNAASVLTLT--NAVLTGAVDNTTGGDNV 339
Db 647 IGNTNALATVNVGAGIATLEGAVIKATTTKLTNAASVLTLTNNAVLTAIDNTTGVNDV 706
Qy 340 GVNLSGALSOVGTGNTGNNTSLATINIGAGVATLDGAVIKATTTKLTDDASVLTFTNPV 399
Db 707 GVLNLSGALSOVGTGNTGNNTSLATISVGAGKATLGGAVIKATTTKLTNDASAVTFTNPV 766
Qy 400 VTGAIDNTGNANNGIVFTGDSVTGNTGNNTNALATISVGAGKATLGGAIKATTTKLTNDAS 459
Db 767 VTGAIDNTGNANNGIVFTGDSVTGNTGNNTNALATISVGAGKATLGGAIKATTTKLTNDAS 826
Qy 460 NAS-----
Db 827 NASAVTFTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGNNTNALATISVGAGKATLGG 886
Qy 463 -----VVTFTGDSVTGSGIGTGLFAT 484
Db 887 ALIKATTTKLTNDASAVTFTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGNNTNALAT 946
Qy 485 VNIGAGITLRAGSLAANNIDFGAASNLEFNGP--AGKN---YNLIGTIANGNNATLNIN 539
Db 947 VNVGAGVTLQAGSLDANNIDFGAASNLEFNGP--AGKN---YNLIGTIANGNNATLNIN 1006
Qy 540 AAGTVIANDVSGIVTVAQINONKIFVINAKNADVDILDAQAISFKGAASRLFLANVS 597
Db 1007 TK-LLTAYHETIGTVAEINIGAGNLEFIDASAGDVITLNAQDIIHFRALDSALVLSNLT 1063
RESULT 2
A41477
190K surface antigen precursor - Rickettsia rickettsii
C:Species: Rickettsia rickettsii
C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 08-Oct-1999
C:Accession: A41477
R:Anderson, B.E.; McDonald, G.A.; Jones, D.C.; Regnery, R.L.
Infect. Immun. 58, 2760-2769, 1990
A:Title: A protective protein antigen of Rickettsia rickettsii has tandemly repeated, ne
A:Reference number: A41477; MUID:90354033; PMID:2117568
A:Accession: A41477
A:Molecule type: DNA
A:Residues: 1-2249 <AND>
A:Cross-references: GB:W31227; NID:gl52465; PIDN:AAA26380.1; PID:gl52466
A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for residue
F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>
Query Match 49.58; Score 1494.5; DB 2; Length 2249;
Best Local Similarity 51.18; Pred. No. 4.2e-66;
Matches 364; Conservative 58; Mismatches 123; Indels 167; Gaps 18;
Qy 20 ALFTSTAAIMLTGS-----GVL---GAARTVTAD--GAELAGTNGIPG---A 60
Db 613 SVLTLTNNAVLTAIDNTTGGDNVGVNLNLSGALSOVGTGNTGNNTSLATISVGAGTATLG 672
Qy 61 GAFVAGSTLQVTAFTV---TDADVSVRALD--LNNFAAGLFSVTGDISLGSVVDVTTGGANK 116
Db 673 GAVIKATTTKLTNAASAVKFTNPVVVTGAIDSTGNANNGIVFTGDSVTGNTGNNTNALATIS 730

Qy 117 LA-VNIDDLGLTLTGTCTAAYGANPALLFOGGGAAAANNVTYALGN-----ITLGGANAGL 171
Db 731 LATNV-----GAGTATLGG-----AVIKATTTKLTNAASVLTLTNNAVL 771
Qy 172 TIAD-----PDVLPITLAGNIDGGIITDNTDAINGTNTGNTPAAQISIGASTLSLGG 227
Db 772 TGAIDNTTGGDNVGVNLNLSGAL-----SOVTDIGTNTSLATISVGAGTATLGG 820
Qy 228 AVIKATTTKLTNAAPVLTLTNNAVLTAIDNTTGGDDVGVNLNLSGALSOVGTGNTGNNTSL 287
Db 821 AVIKATTTKLTNAASVLTLTNNAVLTAIDNTTGGDNVGVNLNLSGALSOVGTGNTGNNTSL 880
Qy 288 LATISVGAGTATLGGAVIKATTTKLTNAASVLTLT--NAVLTGAVDNTTGGDNVGVNLS 345
Db 881 LATISVGAGTATLGGAVIKATTTKLTNAASVLTLTNNAVLTAIDNTTGGDNVGVNLN 940
Qy 346 GALSOVGTGNTGNNTSLATINIGAGVATLDGAVIKATTTKLTDDASVLTFTNPVVVTGAID 405
Db 941 GALSOVGTGNTGNNTSLATISVGAGTATLGGAVIKATTTKLTDAASAVKFTNPVVVTGAID 1000
Qy 406 NTGNANKGVVIFTGASTVTDNIGNTAVLAESVSGAGLLQIOGGVVKANAILTNDAS--- 462
Db 1001 NTGNANNGIVFTGDSVTGNTGNNTNALATISVGAGLLQVGGVVKANTINLTNDASAVT 1060
Qy 463 -----
Db 1061 FTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGNNTNALATISVGAGLLQVGGVVKAN 1120
Qy 463 -----VVTFTGDSVTGSGIGTGLFATVNICAG 490
Db 1121 TINLTNDASAVTFTNPVVVTGAIDNTGNANNGIVFTGDSVTGNTGNNTNALATISVGAG 1180
Qy 491 ITRAGGSLAANNIDFGAASNLEFNGP--AGK--NYNLIGTIANGNNATLNINAAFTVI 545
Db 1181 ITRAGGSLAANNIDFGAASNLEFNGP--AGK--NYNLIGTIANGNNATLNINAAFTVI 1239
Qy 546 ANDVSGIVTVAQINONKIFVINAKNADVDILDAQAISFKGAASRLFLANVS 597
Db 1240 ASHUTIGTVAEINIGAGNLEFIDASAGDVITLNAQDIIHFRALDSALVLSNLT 1291
RESULT 3
T30852
outer membrane protein A - Rickettsia conorii (fragment)
C:Species: Rickettsia conorii
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30852
R:Gilmore Jr., R.D.
Gene 125, 97-102, 1993
A:Title: Comparison of the ompA gene repeat regions of Rickettsiae reveals species-s
A:Reference number: 220904; MUID:93194085; PMID:7690636
A:Accession: T30852
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1029 <GIL>
A:Cross-references: EMBL:L01462; NID:gl52485; PID:gl52486; PIDN:AAA99908.1
C:Genetics:
A:Gene: ompA
Query Match 36.58; Score 1103; DB 2; Length 1029;
Best Local Similarity 42.38; Pred. No. 3.4e-47;
Matches 283; Conservative 60; Mismatches 194; Indels 132; Gaps 18;
Qy 16 GLKTALETTSTAAILMT--GSGVLGAARTVTADGAELAAGTNGIPGAGAF-----VAGSTL 69
Db 365 GNTNALATISVGAGKATLGGAIKATTTKLTNDASVQVTFNPVVVTGAIDNTGNANNGIV 424
Qy 70 QYGAFTVTDADVSVRALDLNNFAAGLFSVTGD-----ISIGS 107
Db 425 TFGTGGSTVGTGNTGNNTNALATISVGAGLLRQGGVVKNSNTINLTNDASVQVTFNPVVVTGA 484
Qy 108 VVDVGGANKLAVNIDDDGLTLT-----LTGTGAAYG----- 138

Db 485 IDNTGNANNGIVTFTGDSVTGNTGNATATISVCGAGKATLGGAIIKATTTKLTNDASQ 544
QY 139 ---ANPALL-----FOGQAAANN--TYTALGNITLGGANAGL----- 171
Db 545 VTFTNPVVVTGAIDNTGNANNGIVTFTGDSVTGNTGNATATVNVG---AGLLRVOGG 601
QY 172 -----TI-----ASDPDLVGPITLAGNID-----GGGIITDNTDAIINGTINTNPAQAI 216
Db 602 VYKSNITNLNDASQVTFNPNVVTGAIDNTGNANNGIVTFTGDSVTGNTGNATATV 561
QY 217 STGASTLSGGAVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGDDVGVNLNGALS 276
Db 662 NVGAGLLRVOGGVKSNTINLNDASQVTFNPNVVTGAIDNTGNANNGIVTFTGN-S 718
QY 277 QYTNIGNTNSLATISVAGTATLGAVIKATTTKLTNAASVLTLT--NAVLTGAVDNTT 334
Db 719 TVTGNIGNTNAATVNVGAGIATLGAVIKATTTKLTNAASVLTLTNNAVLTGAIDNTT 778
QY 335 GGDNVGVNLGALSQVTVTGNIGNTNSLATINIGAGVATLDGAVIKATTTKLTDDASVLIF 394
Db 779 GVDNVGVNLGALSQVTVTGNIGNTNAATISVCGAGKATLGGAIVIKATTTKLTNDASQVTF 838
QY 395 TNPVVVTGAIDNTGNANNGVIFTGASTVTDNIGNTAVLAESVCGAGLLQIOGGVVKANA 454
Db 839 TNPVVVTGAIDNTGNANNGIATFTGDSVTGNTGNATATVNVGAGLLRVOGGVKSNT 898
QY 455 INLNDAS-----VVFTGDSVTGSGTGFELFATVNIAGI 491
Db 899 INLNDASAVTFTNPVVVTGAIDNTGNANNGIVTFTGDSVTGNTGNATATISVCGAG 958
QY 492 TLRAGGSLAANNIDF--GAASNLEFNGPAGKNYLNIGTIANGNNATLINAAQTIVIANDVS 550
Db 959 AILGGAIKATTTKLTNDASQVTFNPNVVT--GAIDNTGNANNGIVTFTGDSVTGNTGN 1017
QY 551 IGTVAQINI 559
Db 1018 TNALATVNV 1026

RESULT 4

B71704

cell surface antigen (sca3) RP451 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii

C>Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000

C:Accession: B71704

R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sierhertz-Ponten, T.; Alsmark, D.

Nature 396, 133-140, 1998

A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A:Reference number: A71630; MUID:99039499; PMID:9823893

A:Accession: B71704

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2340 <AND>

A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAAL4908.1; PID:g386100

A:Experimental source: strain Madrid E

C:Genetics:

A:Gene: sca3; RP451

Query Match

Best Local Similarity 15.0%; Score 452.5; DB 2; Length 2340;

Matches 194; Conservative 84; Mismatches 232; Indels 171; Gaps 34;

QY 26 TAAIMLTGSGVLGAARTVTDAGLAAGTNIIGPAGAFVAGSTPLOTGTAFTVTD---ADV 82

Db 407 TDCVTENNNGVGGTLLIINAKNTISAKLLNATKAKTQIINLTMNPSAGDISIDIAON 466

QY 83 SVRALDNNFAAGLSFVSDGISL---GSDVDTGGKLANVINDDGLTLTLTGTAAYGA 139

Db 467 TIYIDAKN-----GNVLLNNNAKIIIEGADSMALALINTGYTADRTFI----- 510

QY 140 NPALLFOGQAAANNYYTALNTLTGGANAGLTIASDPVLGPTIAGN-----ID 190

Db 511 -----IYNLNLQSGNDEY---GIVKBAIKKVITIANQS---GPYITGQDNTHLRLKELIVE 560

QY 191 GGG-IITDNTDAAINGTICNTNPAAQISIGASTLSLGG--V-KATTTKLTNAAPVLT 246
Db 561 GAGDIIIDT-----IFTKLLSINSTGTQTFN-RTLDLGGAGNTAFKKGHFLVNG----- 610
QY 247 TNANAVLTGAVDNTTGGDDVGVNLNGALSQVTVTGNIG-NTNSLATISVAGTATLGAVI 305
Db 611 -----VTGSI--TWSENNQGITLINS--GNITGVIGTNELGLKLVNIGADPVTCSANVF 660
QY 306 KATTTKLTNAASVLTLTNAV-LTGAVDNTTGGDNVGVNLSGALSQVTVTGNIG-NTNSLAT 363
Db 661 --ASVALTNPPSVLILADGVTLTGEV--TTHNNTKGVLSL-GTGSNITGOIGTNSAALEK 715
QY 364 INIGAGVATLDGAVIKATTTKLTDDASVLIFTNPVVVTGAIDNTGNANNGVVIETGASTV 423
Db 716 INIGAGASNID--SNYAGSTVLTQDTSELTSLLNDVVVNSNIITTAGNNSGKLIETGNGOI 774
QY 424 TDNIG-NTAVLAEV-----GITLRAGGSLAANNIDFGA--ASNLEFNGPAGKN 521
Db 775 TGNIGANGAALOEVEVFNFTNIGGTANSQNFVAHSAANVITGLTTGALKYKDTGTTIA 834
QY 446 QGVV-----KANAINLTDNASV-----VTFTGDSVTGSGIG--GTFLFA 483
Db 835 HGGLVGDIDFNKAGKFFILGDGAMIDGSLCNGGVAGTLDFFIGDGNVTQNIQADNANST 894
QY 484 TVNIGA-----GITLRAGGSLAANNIDFGA--ASNLEFNGPAGKN 521
Db 895 TINOQDNTKNTVIANDIFVDNIHFTNGGILQLOGNLTTHNIDFGANGTLEFNG--NNT 952
QY 522 YNLIGTIANGNNATLINAAQTIVIANDVSIGTVAQINI-----QNNKIFVNAKNADVD 575
Db 953 YNLNAILVNGQGL--NAFTNLKASDDTIGTVKIINIQTGPQN---FTIQVNNKNLT 1007
QY 576 ILDA--QAISFKGAASRLFLA 594
Db 1008 LVSSVNSSINFGDANSOLILS 1028

RESULT 5

JCI340

outer membrane protein B precursor - Rickettsia japonica

C:Species: Rickettsia japonica

C>Date: 30-Sep-1993 #sequence_revision 21-Jan-1997 #text_change 13-Nov-1998

C:Accession: JCI340

R:Yan, Y.; Uchida, T.

Chinese J. Microbiol. Immunol. 16, 220-226, 1996

A:Title: Cloning and sequence analysis of the gene encoding the 120kDa outer membrane

A:Reference number: JCI340

A:Accession: JCI340

A:Molecule type: DNA

A:Residues: 1-1651 <YAN>

A>Note: the authors translated the codon GGT for residue 241 as Cys, CAA for residue

C:Genetics:

A:Gene: ompB

C:Keywords: membrane protein

F:1-34/Domain: signal sequence #status predicted <SIG>

F:35-1651/Product: outer membrane protein B #status predicted <MAT>

Query Match

Best Local Similarity 12.3%; Score 372.5; DB 2; Length 1651;

Matches 185; Conservative 103; Mismatches 260; Indels 213; Gaps 39;

QY 13 IQKGLKTAFTTSTAAMLTGSG-VLGA--RT-----VTADGA---ELAAGTNIIGPG 59

Db 8 LKRLISAGLVTAFTATVAFSGAMGAALQQNRTKVAAVTVDGVGFDOTAPANVAVA 67

QY 60 AGAFVAGSTLOXTGAFTVTDADVSVALDINNFAAGLSFVSDGISLGSVVD-TGGANKLA 118

Db 68 PNAVI---TANANNGLNLTNPAGSFNGLFLSNANNLAVTVSEDTLGLFINNAANNANCFN 124

QY 119 VNIDGGLTTLTGTG-----TAAAGA-NPALLFOGQAAANNYYTALGNITLGGANAGL 171

Db 125 LTLNAGKTLTITGOGITNVQSAATHNAQNIQVAFNGGAIANNDLSGLGTIDFGAAASTL 184

Query Match	11.0%;	Score 333.5;	DB 2;	Length 1026;
Best Local Similarity	24.7%;	Pred. No. 2.7e-09;		
Matches 181;	Conservative 76;	Mismatches 232;	Indels 243;	Gaps
QY	16	GLKTALETTTAAIMLTGSGVL-----GAARTVTAD-GAELAGT-----NTGP	58	
Db	315	GIETNMNVTSGAAITNTSSGVTGLTALINTSGAAQTVTAGAGQLNTATTAAQAANNVAV	374	
QY	59	GAGAFV-----AGSTLQYTGATVTVDVSVRALDLNNEAAGLFSVGTGDSL-----	105	
Db	375	DGAGNVTVASTGVTSGTTF--VCANSAAGTVSVSVANSSTTTTGATVGGTAVTVAQT	432	

Query Match	11.0%;	Score 333.5;	DB 2;	Length 1026;
Best Local Similarity	24.7%;	Pred. No. 2.7e-09;		
Matches 181;	Conservative 76;	Mismatches 232;	Indels 243;	Gaps
QY	16	GLKTALETTTAAIMLTGSGVL-----GAARTVTAD-GAELAGT-----NTGP	58	
Db	315	GIETNMNVTSGAAITNTSSGVTGLTALINTSGAAQTVTAGAGQLNTATTAAQAANNVAV	374	
QY	59	GAGAFV-----AGSTLQYTGATVTVDVSVRALDLNNEAAGLFSVGTGDSL-----	105	
Db	375	DGAGNVTVASTGVTSGTTF--VCANSAAGTVSVSVANSSTTTTGATVGGTAVTVAQT	432	

16	GLKTALETTSTAAILMTGSGVL-----GAARTVTAD-GAELAAGT-----	NI GP 58
QY		
Db		
315	GIETNNVTSGAAILNTSSVGTGLTALNTNWSGAAGTATTAGAQMLTATTAAQAANNVAV	374
QY		
Db		
59	GAGAFV-----AGSTLQYTGATVTVDVSVRALDNLNFAAGLFSVSGDLSL-----	105
QY		
Db		
375	DGGAENVTVASTGVTGTTT--VCANSAAGSVTSVANSSTTTTGAIATVGGTAVTAQT	432
QY		
Db		

```

Query Match      11.0%;   Score 333.5;  DB 2;   Length 1073;
Best Local Similarity 24.7%;   Pred.No. 2.8e-09;
Matches 181; Conservative 76; Mismatches 232; Indels 243; Gaps

QY 16 GLKTALETTTAAIMLTGSGVL-----GAARTVTAD-GAEAAAGT-----NIGP 58
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 362 GIETMMNTSGAATLTNTSGVTLALNTNTSGAQTVTAGQONLTATTAAQANNAV 421

```


RESULT 9

hypothetical protein rompB [imported] - *Rickettsia conorii* (strain Malish 7)
C:Species: *Rickettsia conorii*
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: E97835
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in *Rickettsia conorii* and *Rickettsia prowazekii*.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: E97835
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1655 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03623.1; PID:gl5620207; GSPDB:GN00173
C:Genetics:
A:Gene: rompB

Query Match 11.0%; Score 331.5; DB 2; Length 1655;
Best Local Similarity 24.4%; Pred. NO. 5.6e-09;
Matches 177; Conservative 81; Mismatches 240; Indels 227; Gaps 40;

QY	13	IQKGLTALPTTSTAAMLTGSG-VLGAA-----RT-----VTADGA-----ELAGTNGIBG	59
Db	8	KKLISAGLVTAATAIVASFAGSAAIQOVRTTNAVATTVDGVGFQDTAVPANVAVP	67
QY	60	AGAFVA-----GSTL-----QYTGAF--TWTDADVSVRALDLNNFAAGLFSVTGDSIGLS	107
Db	68	LNAVITAGVNGKITLNPATPAGSFNGLELNTANNLDDVIRE-----DITLGF	112
QY	108	VVD--TGGAANKLAVNIIDGLTTLTGTG-----TAAYGANPALL--FOGQAAANTTYAL	159
Db	113	ITNVVNAHNHFLMLNAGKTLTITGOGITNVQAAATKNANNVVAQVNNNGAAIDNNDOGV	172
QY	160	GNITLGGANAGL--TIASDPDVLGPITLAGNI-----	189
Db	173	GRIDCGAAASTLVFNLANPTOKAPILLGDNAVIVNGANGTLNVNGFIKVSXKSFATVN	232
QY	190	-----DGGGIITDQA-----AINGTIGTNPAAOISTAGS-----TL	223
Db	233	VINIGDQGIIME--NTDADNVNTNLNQAQATITFNGTDG--TGRVLVLLSKNAATDFNVTG	290
QY	224	SLGG---AVIKATTTKLITNAAPVLTLTNANAVLTGAVDNTGG-----DDVGLV	269
Db	291	SLGSLKGIIEFNTVAVNGQ-----LKANAGANAARVICTNGAGRAAGFVVSVDNGKVA	344
QY	270	NLNG-----ALSQVGTGNTGNTSLATISVGA--GTATLGGAVIKATTTKLTN-----	314
Db	345	TIDGQVYAKDMVIOQSANAVGQVNFRIHVVDGTDGTTAFKTAASKVAITQNSFNFTGDFGN	404
QY	315	--AASVLTLTNAVLTGAVDNTG-----GDNVGVW--NLSCALSOVT--CNIGNTNSLATI	364
Db	405	LAQAIIVPNTWTLNG---NFTGDASPNPGTAGVIYTFDANGTLASASADANVAVNTITAI	461
QY	365	NI--GAGVATLDGAVIKATTTKLTDDASVLIFNPVVVTGAIDNT-----GNANKGVVIETG	419
Db	462	EASGAGVQVLLSGT--HAAEELRLCNGAGSVFKADGTVINGKVNOTALVVGALAAAGTITLDD	519
QY	420	ASTVTNIGNTAVLAEVSVSGAGLLQIQGGVVKANALINDNASVYFTGDSIVTGSIGT	479
Db	520	SATITGDIGNA-----GGAAALQGITLANDATKTLTJLGGANIIGANGG--	562
QY	480	ELFATVINIGAGITLRAGGSL-----AANNTI--DFGA-----ASNLFEFG	516
Db	563	-----TINFOAN-----GGTIKLTSTQNNIVWDFDLAIALATDQTGVVDASSLTNAOITLNG	613
QY	517	PACKKNYNLCTTANGNATL--NINAAQTVLA-----NDVSIGTVQAQINTQNKIFVI	567
Db	614	K-----IGTV--GANNKTLGOFNIGSSKTVISDGDVAINELVIGNNGAVQFAHNTYLT	665
QY	568	NAKNA	572

Db 666 RTNA 670

RESULT 10
AD0123

probable autotransporter protein yafH [imported] - *Yersinia pestis* (strain CO92)
C:Species: *Yersinia pestis*
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AD0123
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
ill, M.; Rutherford, K.; Simmonds, M.; Skellon, J.; Stevens, K.; Whitehead, S.; Barrer
Nature 413, 523-527, 2001
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A:Reference number: AB00001, PMID:21470413, PMID:11586360

A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-3705 <KUR>
A/Cross-references: GB:A
C/Genetics:
A/Gene: yaph

Query Match	10.9%	Score	329.5	DB	2	Length	3705
Best Local Similarity	24.8%	Pred. No.	1.7e-08				
Matches	165	Conservative	88	Mismatches	23	Indels	179
						Gaps	33

Qy **Tt** TTAATMLTSG-----VLGAARTVTTADGAELAAGTNIGP---CAGAFAVGST 68
|| || :| |
|| || :| |

Db 1144 TTVTAGELILGANGAFGQTSLLDIASGASANINGYSQTVGAVTNVGTVTLGSGGVLTSGL 12033

Qy 69 LQYTGAFVTTDADVSVRALDLNFAAGLFSVTGDISLGSVVDGTGGANKLAV----- 119
| | : | | | : | : | : | : |
Dd 1204 LTNGGILDTGG-----ALNLTAGGSATVAGGLTGCAGTNINCGN-LSVSAANSGLSG 1255

0: 130 - NITDCIETIETIETCECTAAVCANBAYEACCCCAAANNTEVTYCNTEYCCANACIETIASED 177

Db 1256 QTHADVASVTLTDTGTLG TSA-----VEVLGTNLNLNGANAAMTNVL SG 1299

QY 178 D----VLGPITLAGNIDGGGIITDNTDAAINGTIGNTPAAQISIGASTLSLGGAVIKAT 233

Db 1300 DGTINTNAAVTLSGNNSFSGAHGIGTDGEL--TVGQ---ASNLGASSATVNLG-----TLT 13500

Qy	234	T	K	R	E	N	A	A	P	V	E	T	L	T	N	A	N	A	V	L	T	G	A	V	D	N	T	T	G	G	D	D	V	G	L	N	G	A	L	S	V	-----	TGN	281											
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:																
Dh	1251	S	H	I	T	I	N	O	V	E	R	S	T	A	N	V	I	C	A	V	A	C	C	M	I	T	T	C	A	D	M	A	I	T	A	N	N	C	E	E	A	V	A	I	A	C	N	C	I	M	A	C	N	I	1410

QY 282 IGTNSLA-----TISVGAGTATLGGAVIKATTTKLTNAAASVLTLT-----NAVLTGA 329

```
Db      1411 LGASSVALAGAGDTLSLGGFNGFTGNSVTCGVLTQTDAAEV-TLTSSNGSVNAVYIDI 1469
```

QY 330 VDNTTGGDNVGVVNLGALSQVTG---NIGTNSLATINIG-----AGVATL----- 373

Db 1470 ADATLNLDIALEN-----HVLTCNGLLNVAKNDASTAFDFGSIVGGAFSGIVNLNTTF 1524

QY	3/4	-----DGAVIKATIKLIDDRASVLIFTNFVVVVGALDNTGNA-----NKGVVVIFIGAS-----	421
DP	1525	-----ALSADNAAALARATLKLSDPS-----VTVGATDRTLHGIDLNGGTLIFDGSPPPO	1576

QY 422 -----T VTDNIGNTAVLA EVSVGAG-----LLQIQGGVV--KANAINLTD 459

Db 1577 ANGVVTVTDLALNSGTIS--ITGAGNWNENEHPVTPPNVSLLEQDRGDILLELINAANVTG 1634

QY 460 NASVFTGDSVTGSGTGTELFATVNICAGITLRAGCSLAANNI-DFGAASNLENGPA 518
||: : | | | | | | : : ||| | | | : : | |

Ov	519	GK----	NYNL--	IGTIANGNNATI	NINAASTVIANDV-----	SIGTVAQINTONKKIEV	566
Dd	1635	NNNLDLLVDGTAITS-	-GTQ-----	GVESAIQGGGSIVANA	IHNHYGLTSS-----	NGNG	1688

Db 1683 GSGIYVNYTSALELLANGCANALLATESG-LTANRVLNAEFGVGLV-VDAQNCALTL 1740

QY 567 INAKN 571

1
2
3
4


```

QY 509 ASNLEFNPGAKNYNLIGTIANGNATLNINAAGT-----VIANDVSI 551
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 567 IGTLTSTGNA-----TFASGSYIAVEIDADGSSDRLAVTGTTIANDVSL 611

```

RESULT 13

H98323
hypothetical protein AGR_L_3085 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: H98323
R:Goodner, B.; Hinkie, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: H98323
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1341 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK90114.1; PID:gr15160106; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_3085
A:Map position: linear chromosome

Query Match	10.7%	Score	323.5;	DB 2;	Length	1341;
Best Local Similarity	25.0%;	Pred. No.	1.1e-08;			
Matches 163;	Conservative	90;	Mismatches	195;	Indels	205;
			Gaps	36;		

QY	20	ALFTTAAIMLTGSGVLG----	AARTVTADGAELAAG-TNIG-----	PG-----	59	
Db	332	AKLWEEAAAVTSGEAIIGHRSASEATVTGDGSKWTTGDLQVGDDTDPGGLAGNCTLAV	391			
QY	60	-ACAFVAGSTLOYTGAF-----	TVTADVSVPALDLNPF-----	AAGLSFVSTGDISLGS	107	
Db	392	TAGGSV-DSTVAHLGVVAGATGSAIVDVGKGSVWTVDRNSLEVGSGAGSLAVTG-----	GG	446		
QY	108	VVD-----	TGGANKLAVNIDDDGLTLTGTGTAAAYGANPALIFCGGQAAANNITYFA	158		
Db	447	LVDAAANIIGTNTGGNSVRVSGADSTVKRSRDLNVGLYNGSMTVEAGGAVKSRDGYVA	506			
QY	159	LGNITGGANAGLTIASDPVLGPITLAGNIDGGGIITONTDAAINGT--	IGNTNPAAOI	216		
Db	507	-----	TGGSTSAVTVG-----	DGSSWAMTGTFFVGYASGAT--	539	
QY	217	SIGASTLSLGGAV--	IKATTKTLTNAAPVLTLTNANAVLTGADVNTT-----	GGDDYGVCLNL	271	
Db	540	--GNVTVSNNGAIRATCVTLGLDLAGASGTWTTIGAGSKVTA	VDNCTVNSGVSVDVGFQG--	596		
QY	272	NGALSOVTGN-----	IGNTNSLATISVSG-----	AGTATL-----	GGA	303
Db	597	SGSLSVYNGGSLDAYNLYYVGNALGSSGAVLVSGVSHSVSDGLMWYVGNAGNSVEITGGA	656			
QY	304	VIKATTKL-TNAASVLTLT-----	NAVLTGADVNTT-----	GGDNVGVV-----	342	
Db	657	SLAAPILITATEAGSTGVLSIGAGSGQTARSAGVEARA	IAFCAGNGSTVFNHSETGYTL	716		
QY	343	--NLSCALSOVTGNIGNTNSLATINIGAGVATLDGAVIKATTK-----	LTDASVLI-----	393		
Db	717	SADISGA-GRVVAEAGVT--	TLGNSNSYSGCTTISAGMLKGTAKTSGSGGIVVNAELVDVG	774		
QY	394	---FTPNPVVVTGAIDNTGNANKGVIFTGASTVTDNIGNTAVLA	EVSVGAGLLOIQGGVV	450		
Db	775	GGTLSNAISGTSGFEKTGDCN--	LILLTGNSTYS--	GATA-----	VYAGKLSVNGSL--	821
QY	451	KANAINLTDNASVVTFTGDSIV--	TGSIIGGTFLFATVNIIGAGITILRAGGSLAANNIDPGA	508		
Db	822	-----	AGASVSGSGATVGGTGTIG-----	GLTVNSGGTLAGPN--	855	

Qy	509	ASNLEFNGPAGKYNLIGTIANGNATLNINAACT-----VIANDVSI	551
Db	856	IGTLTSTGNA-----TFASGSTYAVEIDADGSSDRLAVGTITIANDVSL	900

RESULT 14
E97857

cell surface antigen homolog RCL261 [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: E97857
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.;
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; PMID:21442074; PMID:11557893

A;Accession: E97857
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-652 <KUR>
A;Cross-references: GB:
C;Genetics:
A;Gene: RC1261

Query Match	10.3%	Score 310.5;	DB 2;	Length 652;
Best Local Similarity	24.3%;	Pred. NO. 2.2e-08;		
Matches 166;	Conservative	80;	Mismatches 237;	Indels 199;
				Gaps 34;

[illegible]

RESULT 15

G97778
 cell surface antigen homolog [imported] - Rickettsia conorii (strain Malish 7)
 C:Species: Rickettsia conorii
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
 C:Accession: G97778
 R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
 Science 293, 2093-2098, 2001
 A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A:Reference number: A97700; MUID:2142074; PMID:11557893
 A:Accession: G97778
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-256 <KUR>
 A:Cross-references: GB:AE006914; PIDN:AAL03169.1; PID:g15619717; GSPDB:GN00173
 C:Genetics:
 A:Gene: RC0631

Query Match 10.2%; Score 308.5; DB 2; Length 256;
 Best Local Similarity 31.8%; Pred. NO. 1e-08;
 Matches 94; Conservative 43; Mismatches 104; Indels 55; Gaps 13;

QY 294 GAGTATLGGAVIKATTKLTNAASVLTLTN---AVLTGAVDNTTGGDNVGVVNLNLS---CA 347
 DB 3 GAGNIT-----INAAINLTNNNSILTLFDCNITTLTGNNINNTAGVGGGILNLAHDLGS 57

QY 348 LSQVTGNIQNTNSLATINIGAGVATLDGAVIKATTTKLTDDASVLIFTNPVVVTGAIDNT 407
 DB 58 SNIITGDIQNGSLAAVNVLLGAATLSTILKATNINQSNVTLNLDLDDITVTGNIDGA 117

QY 408 GNANKGVIFTGASTVDNIGHTAVLAESVVGAGLQIQGGVVKANAINLTDNASV--VT 465
 DB 118 KGVNGN---FIGNAILNGNIN-----FNILQNGG--NGKILDQSNNTVNSIV 162

QY 466 FTGDSVTGSGIGTGFATVNICAGITLR-----AGSLAANNIDFGAASNLEFNGPAGKN 521
 DB 163 FADSVLAAGTISVNSL---LDVG-GITFNNSNASGGTLLIINTED-----T 203

QY 522 YNLIGTIANGNNATLINAAGTVIANDVSIGTVQAQINQNNKIFVINAKNADVDIL 577
 DB 204 INI--ALLNAIQAKIQINANLTI--NDPSAGDIGDIRIADNTTYTIDRANGNVNLL 255

Search completed: April 14, 2003, 16:11:37
 Job time : 33 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:01:34 ; Search time 28 Seconds
(without alignments)

916.923 Million cell updates/sec

Title: US-09-800-065-2

Perfect score: 3022

Sequence: 1 MANISLKLQRAIQKLGKTA.....MIELSLKIIYPVLLTVVVS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	1510	50.0	1 OMPA_RICCN	Q52657 rickettsia
2	1494.5	49.5	1 OMPA_RICRI	P15921 rickettsia
3	376.5	12.5	1 OMPB_RICJA	O06653 r outer mem
4	375.5	12.4	1 OMPB_RICRI	Q53047 r outer mem
5	333.5	11.0	1 SLAP_CAUCR	P35828 caulobacter
6	333	11.0	1 OMPB_RICPR	Q53020 r outer mem
7	331.5	11.0	1 OMPB_RICCN	Q9kka3 r outer mem
8	299.5	9.9	1 OMPB_RICTY	P96989 r outer mem
9	276.5	9.1	1 OMPB_RICRI	P14914 rickettsia
10	259.5	8.6	1 YDBA_ECOLI	P33666 escherichia
11	259	8.6	1 BIGA_SALTY	P25927 salmonella
12	258.5	8.6	1 PM10_CHLPN	Q9rb65 chlamydia p
13	251.5	8.3	1 AIDA_ECOLI	Q03155 escherichia
14	250.5	8.3	1 AG43_ECOLI	P39180 escherichia
15	250	8.3	1 SLAP_CAMEE	P35827 campylobact
16	249	8.2	1 PMP6_CHLPN	Q92899 chlamydia p
17	243.5	8.1	1 YPUA_ECOLI	P52143 escherichia
18	238.5	7.9	1 PM20_CHLPN	Q92812 chlamydia p
19	236	7.8	1 PM21_CHLPN	Q92605 chlamydia p
20	229	7.6	1 YF48_MYCTU	Q10778 mycobacteri
21	228	7.5	1 YDBK_ECOLI	P32051 escherichia
22	224.5	7.4	1 PMPB_CHLMU	Q9pjy2 chlamydia m
23	223	7.4	1 Y278_MYCTU	P56877 mycobacteri
24	221	7.3	1 PMP7_CHLPN	Q92898 chlamydia p
25	219.5	7.3	1 YRP5_IRV6	P18309 chilo iride
26	219	7.2	1 PHX5_MOUSE	P08399 mus musculus
27	213	7.0	1 FHAB_BORPE	P12255 bordetella
28	212	7.0	1 VACA_HELPU	Q92kw5 helicobacte
29	210	6.9	1 APMU_PIG	P12021 sus scrofa
30	205	6.8	1 YCGV_ECOLI	P76017 escherichia
31	205	6.8	1 CBPA_CLOCL	P38058 clostridium
32	204.5	6.8	1 PMPB_CHLTP	O84418 chlamydia t
33	204	6.8	1 PMP9_CHLPN	Q9z398 chlamydia p

ALIGNMENTS

RESULT 1

OMPA_RICCN STANDARD; PRT: 2021 AA.
AC Q52657; P95591; P95592; P95593; P95594; Q52667; Q52668; Q52669;
AC Q52670; Q52674;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
DE antigen) (rOmPA) (rOmp A).
GN OMPA OR RCI273.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=94171067; PubMed=8125327;
RA Crocquet-Valdes P.A., Weiss K., Walker D.H.;
RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia
RT conorii (Malish 7 strain).";
RL Gene 140:115-119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [3]
RP SEQUENCE OF 8-204 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RX MEDLINE=97015921; PubMed=8862558;
RA Roux V., Fournier P.E., Raoult D.;
RT "Differentiation of spotted fever group rickettsiae by sequencing and
RT analysis of restriction fragment length polymorphism of PCR-amplified
RT DNA of the gene encoding the protein rOmpA.";
RL J. Clin. Microbiol. 34:2058-2065(1996).
RN [4]
RP SEQUENCE OF 953-2012 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RX MEDLINE=97015921; PubMed=8862558;
RA Raoult D., Fournier P.E., Roux V.;
RT "Phylogenetic analysis of spotted fever group rickettsiae by study
RT of the outer surface protein rOmpA.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -!- PTM: GLYCOSYLATED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

Q8x8v7 escherichia
P55981 helicobacte
P45508 escherichia
P04949 escherichia
Q48258 helicobacte
P16466 proteus mir
O53553 mycobacteri
P13126 deinococcus
Q48245 helicobacte
O53810 mycobacteri
P2835 saccharomyc
Q48247 helicobacte

PI5921;
 01-APR-1990 (Rel. 14, Created)
 01-APR-1990 (Rel. 14, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (rOmpA) (rOmpA).
 DE antigen (rOmpA) (rOmpA).
 OS OMPA.
 GN Rickettsia rickettsii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=783;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R;
 RX MEDLINE=90354033; PubMed=2117568;
 RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
 RT "A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences.";
 RL Infect. Immun. 58:2760-2769(1990).
 CC -!- FUNCTION: ELICITS PROTECTIVE IMMUNITY.
 CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -!- PTM: GLYCOSYLATED (PROBABLE).
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@sib-sib.ch).
 CC
 CC EMBL; M31227; AAA26380.1;
 DR PIR; A41477;
 DR InterPro: IPR003858; rOmpA_rOmpB.
 DR Pfam; PF02708; rOmpA_rOmpB; 1.
 KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 2249 OUTER MEMBRANE PROTEIN A.
 FT DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 212 286 A (TYPE I).
 FT REPEAT 287 358 B (TYPE II).
 FT REPEAT 359 430 C (TYPE II).
 FT REPEAT 431 505 D (TYPE II).
 FT REPEAT 506 577 E (TYPE II).
 FT REPEAT 578 652 F (TYPE II).
 FT REPEAT 653 724 G (TYPE II).
 FT REPEAT 725 799 H (TYPE II).
 FT REPEAT 800 874 I (TYPE II).
 FT REPEAT 875 949 J (TYPE II).
 FT REPEAT 950 1021 K (TYPE II).
 FT REPEAT 1022 1093 L (TYPE II).
 FT REPEAT 1094 1165 M (TYPE II).
 FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
 SQ SEQUENCE 2249 AA; 224333 MW; A9D6646C089DF087 CRC64;
 Query Match 49.5%; Score 1494.5; DB 1; Length 2249;
 Best Local Similarity 51.1%; Pred. No. 2e-61;
 Matches 364; Conservative 58; Mismatches 123; Indels 167; Gaps 18;
 QY 20 ALFTTAAIMLRGS-----GYL---GAARTVTAD--GAELAAAGTNIGPG---A 60
 Db 613 SVLTLTNANAVLRGADTNTGGDNGVNLNGALSGQVTGDTGNLTSLATISVGAGTATLG 672
 QY 61 GAFVAGSTLOYTGAFV---TDADSVRALD-LNNFAAGLFSVTGDISLGSVVDTGANK 116
 Db 673 GAVIKATTTKITNAVSARKFTNPVVTGADISTGNANNGIVTFTGNSVTG--DIGNINA 730
 QY 117 LA-VNIDGGLTLTLTGCTAAAGNALLFQGGGAANNTYTAIGN-----ITLGGANGL 171
 Db 731 LATVNV-----GAGTATLGG-----AVIKATTTKLTNAASVLTLTNANAVL 771

QY 172 TIASD-----PDVLGPITLACNIDGGGIITDNTDAAINGTICNTNPAAQISIGASTLSGG 227
 Db 772 TGAIDNTGGDNGVNLNGAL-----SQVTGDTGNLTSLATISVGAGTATLG 820
 QY 228 AVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTGGDDVGLNGLNGALSGQVTGNTGNS 287
 Db 821 AVIKATTTKLTNAASVLTLTNANAVLTGAVDNTGGDNGVNLNGALSGQVTGNTGNS 880
 QY 288 LATISVGAGTATLGGAVIKATTTKLTNAASVLTLT---NAVLTGAVDNTGGDNGVNVNLS 345
 Db 881 LATISVGAGTATLGGAVIKATTTKLTNAASVLTLTNAASVLTLTNAIDNTTGGDNGVNLN 940
 QY 346 GALSQVTGNTGNTSLATINIGAVATLDCGAVIKATTTKLTDDASVLIETNPVVVTGAID 405
 Db 941 GALSQVTGDTGNLTSLATISVGAGTATLGGAVIKATTTKLTDRASVKTNPVVVTGAID 1000
 QY 406 NTGNANKGVVIFTGASTVTDNIGNTAVLAESVSGAGLQIQGGVVKANAILNTDNAS--- 462
 Db 1001 NTGNANNGIVTFTGNSVTGNVGTNALTAVNVGAGLQVGGVVKANTILNTDNASAVT 1060
 QY 463 ----- 462
 Db 1061 FTNPVVVTGADTNGNANNGIVTFTGNSVTGNVGTNALTAVNVGAGLQVGGVVKAN 1120
 QY 463 -----VVTFTGDTGTVTSGISGTELFATVNIGAG 490
 Db 1121 TINTDNASAVTNPVVVTGADTNGNANNGIVTFTGNSVTGDTGNLTALATVNVGAG 1180
 QY 491 ITRAGGSLAANNIDFGAASNLEFNGP---AGK--NYNLGTIANGNNATLNINAAGTVI 545
 Db 1181 ITRAGGSLAANNIDFGARSTLEFNGPLDGGKRAIPYFKGAANGNAILNVNTK-LLT 1239
 QY 546 ANDVSIGTVQAQINNTONKIFVINAKNADVDILDAQAISFKGAASRLFLANVS 597
 Db 1240 ASHTLTGTVAEINIGAGNLTFTIDASVGDVITLNAQINFRARDSVLVLSNLT 1291
 RESULT 3
 OMPB_RICJA STANDARD; PRT; 1656 AA.
 ID OMPB_RICJA
 AC O06653;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein) (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOmpB) (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 DE OMPB.
 GN Rickettsia japonica.
 OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YH;
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia japonica.";
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL FACTOR WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY SIMILARITY).
 CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

QY 155 TYTALGNITLGGANAGLTIASDPVLPITLAG-----NIDG 191
 Db 490 TAGIATVTLGSGFAA-TI--DSSALTNVLSGTGTSLGIRGALTATPTANTLTNLVNG 546
 QY 192 ---GGIITDNTDAA-----INGTIGNTPAAQISIGASTLSLGGAVIKATTTKLTNA 240
 Db 547 LTTTGALTDEAAADGFTTINAGSTIASSTIASLVAADATTLNISGARVTTTSHATAA 606
 QY 241 APVLTFLNA-----NAVLTGAVDNT----- 260
 Db 607 LTGITVTSVGCATLGAELATGLVFTGAGADSIILGATTAIVNGAGDDTVTVSSATLGA 666
 QY 261 ---TGCD--DVGVNLNGALSOVQTNIGNTNSLATISVGAGTATLG-----G 302
 Db 667 GGSVNGGDDVLDLVANVNGSFSADPAFGGFETLRV-----AGAAAGSHNANGFTALQLG 722
 QY 303 AVIKATTTKLTNAA-----SVLTTLNAVLTGAVDNTTGGDNVGVNLSGALSQVGTGIGN 357
 Db 723 ATAGATT--FTNAVNVGLVLAAPTGTITVLANATGTSDFENLTSSSAAALAGIVA- 779
 QY 358 TNSLATINIGA-----GVATLDGAVIKATTTKLTDDASVLIFTPVVTGATDNTGNANG 413
 Db 780 LAGVETVNIATDNTTTHAVDTLTQATSAT-----SIVVTG-----NAG 819
 QY 414 VVIETGASTVDNIGNTAVLAESVSGAGLLOIGGVVYKANAINLTDNASVVTFGDSVT 473
 Db 820 LNL-----TNGNTA-----VTSFASAVTGTGSAVTFVSANTTV 854
 QY 474 GSIGGTELFATVNIAGITLRAGSLAANNIDFGA-ASNLEFNGPAGKNYLNIGTIANGN 532
 Db 855 GEV-----VTIRGAGADSILGTSATANDIIGGAGADTLVYTG-----DTFTGGT 901
 QY 533 NATI--NINAAGT-----VIANDVSGTVAQINQNKIFVINAKNADVDILDAQAISPKGA 587
 Db 902 GADIFDINAIGTSTAFVTTIDAAG-----DKLDVLGISTNGA 939
 QY 588 -ASRLFLANVSL 598
 Db 940 IADGAFCAAVTL 951

RESULT 6
 OMPB_RICPR STANDARD; PRT; 1643 AA.
 AC Q3020; Q9ZCM0;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOMPb)
 DE (rOMP B) (Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide).
 DE OMPB OR SPAP OR SPA OR RP704.
 GN Rickettsia prowazekii.
 OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RP STRAIN=Breil;
 RX MEDLINE=91045972; PubMed=2122457;
 RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
 RT "Characterization of the gene encoding the protective paracrystalline-
 surface-layer protein of Rickettsia prowazekii: presence of a
 truncated identical homolog in Rickettsia typhi.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Breil;
 RA Moron C.G., Yu X.J., Walker D.H.;
 RT "Sequence analysis of ompB of Rickettsia prowazekii.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039439; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 RN [4]
 RP PARTIAL SEQUENCE.
 RC STRAIN=Breil;
 RX MEDLINE=92114896; PubMed=1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNB fragments of
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
 RT prowazekii.";
 RL Mol. Immunol. 29:95-105(1992).
 RN [5]
 RP CLEAVAGE SITE.
 RC MEDLINE=92104668; PubMed=1729180;
 RX Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 RT membrane protein of rickettsiae: identification of an avirulent mutant
 RT deficient in processing.";
 RL Infect. Immun. 60:159-165(1992).
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M37647; AAA26390.1; ALT_INIT.
 DR EMBL; AF161079; AAD42234.1; -.
 DR EMBL; AJ235273; CAA15140.1; -.
 DR InterPro; IPR003858; rOmpA_rOmpB.
 DR Pfam; PF02708; rOmpA_rOmpB; 1.
 KW Antigen; S-layer; Cell wall; Complete proteome.
 FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.
 FT VARIANT 257 257 V -> A (IN STRAIN BREIL).
 FT VARIANT 1010 1010 Y -> D (IN STRAIN BREIL).
 FT VARIANT 1450 1450 A -> S (IN STRAIN BREIL).
 FT CONFLICT 178 179 AA -> VC (IN REF. 1).
 FT CONFLICT 191 201 TTQEAFLTGA -> INSRSSSVHLVS (IN REF. 1).
 FT CONFLICT 212 212 T -> I (IN REF. 1).
 FT CONFLICT 313 313 Q -> L (IN REF. 1).
 FT CONFLICT 1104 1104 D -> G (IN REF. 2).
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).
 SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;
 Query Match 11.0%; Score 333; DB 1; Length 1643;
 Best Local Similarity 24.4%; Pred. No. 1.9e-08;
 Matches 187; Conservative 85; Mismatches 272; Indels 222; Gaps 38;
 QY 13 IQKGLTKALFTTSTAATMLTGSV-LGAA---RTVTAD-----GAELAAAGTNGPGA 60
 Db 8 LKKIISAGLVATATIVAGFSGVGAAGAAQYNNRTTNAATTFDGGFDQAAGANI-PVA 66
 QY 61 GAFVAGSTLOYTGAFVTVTDADSVRLDNNFAAGLFSVTGDISLGSVDTGGANK-LAV 119
 Db 67 PNSVIITANNPITFNTPNHNSLFLDTANDLA--VTINEDITLGTITTAQAQKFFNE 124

Db 473 --IHGAELRLNAGSIFKLADCTVINGPVNQLVNNALAAAGSILQDGSALITGIDGNG 530
QY 431 AV---LAESV-----GAGLL-----QIOGGVVKANAIN-----456
Db 531 AVNALQDITLADASKILTLGSGANILIGANAGATHFOANGGTIOLTSTQNILVDFDL 590
QY 457 -LTDNASV---TFTGDSVT---GSGT-----GTFLFATVNIAGITLRAAGGSLAANNIDF 506
Db 591 VTTDQGVVDASSLNNQTLTNGSIGTIGANTKTLGRENVSSTILNAGDVAINE-- 648
QY 507 GAASLNFNGPAGKYN--LICTIANGNATININNA-----GTVIANDVSTGTVAQINI 559
Db 649 -----VWENEDSGVHLTHNTYLIITKTINAAOQKRIIVAADINTDALTALDNLGSAE--SP 702
QY 560 ONNKIFVINAKNADVDILDAQAISFKGAASRLFLANVS 597
Db 703 LSNHFAKKAAGD-----SILHIGKGVNLYANNIT 733
RESULT 9
ID 120K RICRI STANDARD; PRT: 1300 AA.
AC P14914;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 120 kDa surface-exposed protein.
GN P120.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=90136087; Pubmed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kDa surface-exposed protein of Rickettsia rickettsii";
RL Mol. Microbiol. 3:1579-1586(1989).
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS
CC RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -!- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPITOPES
CC CONFERRING ANTIGENICITY TO THE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16353; CAA34402.1; -.
DR PIR: S07575; S07575.
DR InterPro: IPR003858; rOmpA_rOmpB.
DR Pfam: PF02708; rOmpA_rOmpB; 1.
KW Antigen; Glycoprotein; Cell wall; S-layer.
FT CARBOHYD 7 7
FT CARBOHYD 66 66 N-LINKED (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (POTENTIAL).
FT CARBOHYD 436 436 N-LINKED (POTENTIAL).

FT CARBOHYD 444 444 N-LINKED (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (POTENTIAL).
FT CARBOHYD 799 799 N-LINKED (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (POTENTIAL).
FT CARBOHYD 826 826 N-LINKED (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (POTENTIAL).
FT CARBOHYD 861 861 N-LINKED (POTENTIAL).
FT CARBOHYD 879 879 N-LINKED (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (POTENTIAL).
FT CARBOHYD 1116 1116 N-LINKED (POTENTIAL).
FT CARBOHYD 1128 1128 N-LINKED (POTENTIAL).
FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
FT CARBOHYD 1146 1146 N-LINKED (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (POTENTIAL).
SQ SEQUENCE 1300 AA; 132801 MW; E09E52C3F647243D CRC64;
Query Match 9.1%; Score 276.5; DB 1; Length 1300;
Best Local Similarity 25.0%; Pred. No. 5.6e-06;
Matches 138; Conservative 74; Mismatches 209; Indels 131; Gaps 27;
QY 98 SVTGISLGSVVDGTGANKLAVNIDDLTLTLTGCTGAAYGANPALFQGGQAAANTYT 157
Db 7 NATGVNFRHIVDVGA-----DGTAFKT-----AASKVTIT 38
QY 158 ALGNTITLGGANAGLTIASDPDVLGPITLAGNIDGGIITDNTDAAI-----NGTIGNTPA 213
Db 39 QDSN--FGNTDFG-NLAAQIKVPNAITLTGFTGDSNPGNTAGVITFDANGTLESASD 95
QY 214 AQISI-----GASTLSLGGAVIKATTKLTNAAPVLTITNANAVLTGAVDNT-- 261
Db 96 ANVAVTNNTAIEASGAGVQLSGT--HAAELRLNAGSIFKLAD-GTVINGKVNOTALV 152
QY 262 -GGDDVGLNLGALSQVNTGNIGNTNSLATI-----SVGAGTATLGGAVI----- 305
Db 153 GGALAAGTTLDGS-ATITGDIAGNAGGAALQRIITLANDAKKTLTGGANILGAGGGTID 211
QY 306 ---KATTKLTNAASVLTITNNAVLTGAVDNTTGGDNNVNVNLSGALSQVT-----GNI 355
Db 212 LQANGTILKTS-----TQNNIVVD--FDLAIATDQGVVDASSLTNAQTLLINGKIGTI 264
QY 356 G-NTNSLATINICAGVATLDGAVIKATTKLTDDASVLFTNPVVVTGADNTGNANKGV 414
Db 265 GANNKTLGOFNGSSKTVLSNGNVAINELVIGNDGAVQFAHDTYLTIT---RTTNAAGQGGK 321
QY 415 VIF---TGASTVT---DNIGN-TAVLAESVSGAGLQIOGGVVKANAINL-----TDNA 461
Db 322 IIFPVVNGTTLAAGTNLGSATNPALAEINFGSGKGVNVDVTLNVGECVNLATNITTTDA 381
QY 462 SVVTF-----GSTVTVGSIGGTE--LPATVNICAGITLRAAGGSLAAN-NIDFGAASNLEF 514
Db 382 NVGSFVFNAGGTNIVSGTVGGQGNKFNVALENGTITVKFLGNATENGNTTIAANSPLQI 441
QY 515 NGPAGKNYNLIGTIANGNNATINNAAGTVIANDVSTGTVAQINIQNKKIFVINAKNADV 574
Db 442 GG-----NY-----TADCVASADGTGIVEFVNTGPIT-----VTLNKQAAPV 478
QY 575 DILDAQAISFKG 586
Db 479 NALKQITVSGPG 490
RESULT 10
YDBA_ECOLI
ID YDBA_ECOLI STANDARD; PRT: 2003 AA.
AC P33666; P76087; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Hypothetical protein ydbA.
GN	YDBA OR B1401/B1405.
OS	Escherichia coli.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Escherichia.
OX	NCBI_TaxID=562;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=K12 / MG1655;
RA	MEDLINE=97426617; PubMed=9278503;
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA	Mau B., Shao Y.,
RT	"The complete genome sequence of Escherichia coli K-12."
RL	Science 277:1453-1474(1997).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=K12;
RA	MEDLINE=97251357; PubMed=9097039;
RA	Alba H., Baba T., Fujita K., Hayashi K., Inada T., Itoh T.,
RA	Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA	Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA	Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA	Sampel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA	Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.,
RT	"A 570-kb DNA sequence of the Escherichia coli K-12 genome
RL	corresponding to the 28.0-40.1 min region on the linkage map."
RL	DNA Res. 3:363-377(1996).
RN	[3]
RP	SEQUENCE OF 464-2003 FROM N.A.
RC	STRAIN=K12;
RA	MEDLINE=92190338; PubMed=1665988;
RA	Moszer I., Glaser P., Danchin A.,
RT	"Multiple IS insertion sequences near the replication terminus in
RL	Escherichia coli K-12."
RL	Biochimie 73:1361-1374(1991).
CC	-1- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYSG (AC P25928).
CC	-1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
CC	THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
CC	BETWEEN AMINO ACIDS 839 AND 840.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AE000237; AAC74483.1; ALT_SEQ.
DR	EMBL; AE000237; AAC74487.1; ALT_SEQ.
DR	EMBL; D90778; BAA15009.1; ALT_SEQ.
DR	EMBL; D90778; BAA18880.1; ALT_SEQ.
DR	EMBL; D90779; BAA18881.1; ALT_SEQ.
DR	EMBL; X62680; -; NOT_ANNOTATED_CDS.
DR	EcoGene; EG11307; ydbA.
KW	Hypothetical protein; Complete proteome.
FT	CONFLICT 489 489 I -> V (IN REF. 2).
FT	CONFLICT 495 495 I -> V (IN REF. 2).
SQ	SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;
Query Match	8.6%; Score 259.5; DB 1; Length 2003;
Best Local Similarity	22.9%; Pred. No. 5.2e-05;
Matches 179; Conservative	93; Mismatches 257; Indels 251; Gaps 44;
QY	19 TALETTSTRAIMLTGSGVGAARTVADGAELAAAGNIGPGAGVAGSTFLQYTGFTVT 78
DB	925 TGLVQSNSTIINTDSGII-----DLVGRGSGVGLA---TADSTAENOGKITLD 970
QY	79 DADVSVRALDNNFAAGLFSVTGDISIGSVVDG-----GANKLAVNTDGG- 124

	Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K., Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RA	"Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA.";
RT	Nucleic Acids Res. 28:2311-2314(2000).
RN	[4]
RC	SEQUENCE OF 1-914 FROM N.A.
RX	MEDLINE=CWL029/VR-1310;
RA	Knudsen K., MadSEN A.S., Mygind P., Christiansen G., Birkelund S.;
RA	"Identification of two novel genes encoding 97- to 99-kilodalton outer membrane proteins of Chlamydia pneumoniae";
RT	Infect. Immun. 67:375-383(1999).
RL	-1 SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC	(POTENTIAL).-----
CC	-1 SIMILARITY: BELONGS TO THE PMR OUTER MEMBRANE PROTEIN FAMILY.-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licensese@isb-sib.ch) -----
EMBL:	AJ133034; CAB37071.1; -
DR	EMBL: AE002192; AAF38160.1; -
DR	EMLB: AP002546; BA98657.1; -
DR	EMLB: AJ001311; CA004871.1; -
DR	PFI1-2DPAGE; OH6163; -
TIGR:	CP0303; -
InterPro:	IPIR003368; Chlamydia_pmp.
IPRO:	IPR003357; OMP.
Pfam:	PF02385; OMP; 1.
Pfam:	PF02415; DUF145; 2.
KW	Outer membrane; Signal; Multigene family.
FT	SIGNAL 1 25 POTENTIAL.
CHAIN	26 928 PROBABLE OUTER MEMBRANE PROTEIN PMP10.
SEQUENCE	928 AA; 97229 MW; 059D05206AIDDEI CRC64;
Query Match	8.6%; Score 258.5; DB 1; Length 928;
Best Local Similarity	23.9%; Pred. No. 2.7e+05;
Matches	147; Conservative 94; Mismatches 235; Indels 139; Gaps 26;
QY	21 LFTSTAAIMLTGGVGLGAARTVTADGAEALAGNTNTPGAGAVFVGASTLOYTGAFVTDA 80 : : : : : : : : : : : : 8 LVLSITACETSCGTFFAA---TAEE----NIGPSDFSGSTNGTYTKNNITTGI 55
Dd	81 DVSVRALDLNFNAAGFSVTDISLGSDVDPTGCANKLVANNIDGLTLTLTGCTAYGAN 140 : : : : : : : : : : : : 56 D-----YYLTGDITLQLMGDSALTTCGFSS-DTESLSFAKGYSLFN 99
QY	141 PALFOGQGQAANTYTALGNITLGGANAAGLTIASDDVLGPITLAGNIDGGIIIT-DNT 199 : : : : : : : : : : : : 100 ----IKSSABEGALSIVTDKNLSLTFGSLLFLAAPSSVVITPSKGAKCGDGILFDN- 154
QY	200 DAALNCITI-----GWNTPAQAISICASTLSL-----GGAVIKATTTK 236 155 ---NGTLEPKDYCEENGAIISTKLNSLKNTSGSISEFGNKSSATCKRGKAICATGTV 210
Dd	237 LT-AAAPVLTLTNANAVLTGADVNTGGDDGVNLINGALSQTGIGNINTSLA-----T 290 211 ITNNTAPTFLSNSIAEAGAAGINSTCMTITGNTSLVFSENSVTAGNGALLSGDADVT 270
QY	291 IS-----VGACTATLGGVAIKATTKTUNAASUTLNNAVLTGAVDNVGVVN 343 271 ICSNQVSIFSGNOAVANGAIGAIAKYLUKTSAGGGGGISFNSNVIOG---TTAG-NGGAIS 325
QY	344 LSQUALSOVTGNIGNTSNLATINIGAGVATLDGAVIKATTTKTLDASV----- 391 : : : : : : : : : : : : 326 ILAA-----GECSLSAEAADIIFNGNAIVAATPPQTKRNISDIGSTAKITNLR 373
QY	392 -----LIIFTNPVVTGAINDTG--NANK-----GVVIPTG-----ASTVDNI 427

RA Sivasundaram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RA "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN=ML 308-225;
RC Henderson I.R., Owen P.;
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RL [4]
RN PRELIMINARY SEQUENCE OF 53-78.
RP STRAIN=ML 308-225;
RX MEDLINE=89291704; PubMed=2661530;
RA Caffrey P., Owen P.;
RT "Purification and N-terminal sequence of the alpha subunit of antigen
RT 43, a unique protein complex associated with the outer membrane of
RT Escherichia coli.";
RL J. Bacteriol. 171:3634-3640(1989).
RN [5]
RN SEQUENCE OF 53-63.
RP STRAIN=K12 / EMC2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12";
RL Electrophoresis 18:1259-1313(1997).
RN [6]
RN GENE NAME.
RP MEDLINE=97257509; PubMed=9103983;
RX Henderson I.R., Meehan M., Owen P.;
RA "Antigen 43, a phase-variable bipartite outer membrane protein,
RT determines colony morphology and autoaggregation in Escherichia coli
RT K-12.";
RL FEMS Microbiol. Lett. 149:115-120(1997).
CC -1- FUNCTION: CONTROLS COLONY FORM VARIATION AND AUTOAGGREGATION. MAY
CC SUBSTANTIAL AS AN ADHESIN.
CC -1- SUBUNIT: HETERODIMER OF THE PERIPHERAL MEMBRANE PROTEIN (ALPHA
CC CHAIN) ANCHORED TO THE INTEGRAL OUTER MEMBRANE PROTEIN (BETA
CC CHAIN).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE ASSOCIATED.
CC -1- SIMILARITY: TO ADHESIN AIDI-1 AND TO BORDETTELLA PERTACTIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000291; AAC75061.1; ALT_INIT.
CC EMBL; D90838; BAA15825.1; ALT_INIT.
CC EMBL; D90839; BAA15832.1; ALT_INIT.
CC EMBL; U24429; ABA47869.1;
CC HSPSP; P07505; I5RD
CC Ecogene; EG12686; flu.
CC InterPro; IPR004899; Pertact_sup.
CC Pfam; PF03212; Pertactin; 1.
CC Outer membrane; Signal; Complete proteome.
CC SIGNAL 1 52
CC CHAIN 53 551 ANTIGEN 43 ALPHA CHAIN.
CC CHAIN 552 1039 ANTIGEN 43 BETA CHAIN.
CC VARIANT 2 2 K -> N (IN STRAIN ML 308-225).
CC VARIANT 41 42 SL -> FF (IN STRAIN ML 308-225).
CC VARIANT 46 46 T -> K (IN STRAIN ML 308-225).
CC VARIANT 157 157 W -> L (IN STRAIN ML 308-225).
CC VARIANT 188 188 V -> F (IN STRAIN ML 308-225).
CC VARIANT 303 305 ATN -> STI (IN STRAIN ML 308-225).
CC VARIANT 320 320 A -> T (IN STRAIN ML 308-225).
CC VARIANT 372 372 N -> Q (IN STRAIN ML 308-225).
CC VARIANT 493 493 E -> V (IN STRAIN ML 308-225).
CC VARIANT 497 497 S -> N (IN STRAIN ML 308-225).
CC VARIANT 497 497 S -> N (IN STRAIN ML 308-225).
CC

	This SWISS-PROT entry is copyright. It is produced through a collaboration
	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
	the European Bioinformatics Institute. There are no restrictions on its
	use by non-profit institutions as long as its content is in no way
	modified and this statement is not removed. Usage by and for commercial
	entities requires a license agreement (see http://www.isb-sib.ch/announce/
	or send an email to license@isb-sib.ch).

DR	EMBL; AE000291; AAC75061.1; ALT_INIT.
DR	EMBL; D90838; BAA15825.1; ALT_INIT.
DR	EMBL; D90839; BAA15832.1; ALT_INIT.
DR	EMBL; U24429; ABA47869.1; --
DR	HSP; P07505; ISR.D.
DR	EcoGene; EG12686; flu.
DR	InterPro; IPR004895; Pertact_sup.
DR	Pfam; PF03212; Pertactin; 1.
KW	Outer membrane; Signal; Complete proteome.
	SIGNAL 1 52
FT	CHAIN 53 551
FT	CHAIN 552 1039
FT	VARIANT 2
FT	VARIANT 41 42
FT	VARIANT 46 46
FT	VARIANT 157 157
FT	VARIANT 188 188
FT	VARIANT 303 305
FT	VARIANT 320 320
FT	VARIANT 372 372
FT	VARIANT 493 493
FT	VARIANT 497 497
	ANTIGEN 43 ALPHA CHAIN.
	ANTIGEN 43 BETA CHAIN.
	K -> N (IN STRAIN ML 308-225).
	SL -> FF (IN STRAIN ML 308-225).
	T -> K (IN STRAIN ML 308-225).
	W -> L (IN STRAIN ML 308-225).
	V -> F (IN STRAIN ML 308-225).
	ATN -> STI (IN STRAIN ML 308-225).
	A -> T (IN STRAIN ML 308-225).
	N -> Q (IN STRAIN ML 308-225).
	E -> V (IN STRAIN ML 308-225).
	S -> N (IN STRAIN ML 308-225).

FT VARIANT 585 585 H -> Y (IN STRAIN ML 308-225).
FT VARIANT 709 709 E -> K (IN STRAIN ML 308-225).
FT VARIANT 721 721 M -> T (IN STRAIN ML 308-225).
FT VARIANT 753 753 GHL -> SHF (IN STRAIN ML 308-225).
FT VARIANT 803 803 S -> P (IN STRAIN ML 308-225).
FT VARIANT 815 815 A -> V (IN STRAIN ML 308-225).
FT VARIANT 824 824 C -> S (IN STRAIN ML 308-225).
FT VARIANT 835 835 LNLVHTS -> MNLIYNA (IN STRAIN ML 308-225).
FT VARIANT 845 847 OGT -> LGA (IN STRAIN ML 308-225).
FT VARIANT 855 855 S -> T (IN STRAIN ML 308-225).
FT VARIANT 888 888 Q -> L (IN STRAIN ML 308-225).
FT VARIANT 1025 1025 S -> I (IN STRAIN ML 308-225).
FT CONFLICT 61 63 ETV -> TTT (IN REF. 5).
SQ SEQUENCE 1039 AA; 106841 MW; 5170D647C8DEBE0 CRC64;

Query Match 8.3%; Score 250.5; DB 1; Length 1039;
Best Local Similarity 24.2%; Pred. No. 7e-05;
Matches 150; Conservative 76; Mismatches 230; Indels 163; Gaps 30;

QY 33 GSGVLGAARTVTDGAE-----LAAGTNIG-----PGAGAFVAGSTLQYTGFTV 77
Db 139 GQSLQGRAVNTTLNGGBQWHEGAITGTVINDKGWQVVRPGT---VATDTVVNTGAEGG 195

QY 78 TDADVSVRALDNNFAAGLFSVTGDISLGSVDTGGANKLAVNIDDLGLTLTGTGTAA 137
Db 196 PDAB-----NGDTGQF-VRGD-----AVRTINKNGRQIVRAEGT----- 229

QY 138 GANPALLFOGQAAANNYYTALGNITLGG-----ANAGLTIASDPDLGPTLGNIDGG 193
Db 230 -ANTVVYAGDQTVHG--HALDTLNGYQYVHNGG--TASDTVV-----NSDGWQ 276

QY 194 IITDNTDAAINGTIGNT--NPAAQISIGAS-----TSLGGAVIKATTTKLT----- 238
Db 277 IVKN-----GGVAGNTVTKGRLQVDAGTATNVTLKQGGALVSTSTAAVTGTINRLGA 330

QY 239 -----NAAPVLTNLNANA--VLTCADVNTTGGDDVGVNLNGALSQVGTGNIGNTNSLATI 291
Db 331 FSVVEGKADNVVLENGRGLDVLTHGTATNTRVDDGGTLDRVNGGTATTVSMNGGVILLAD 390

QY 292 SVAGAGTAT-----LGGAVIKATTTKLTNAASVL-----TLTNAVLTGADVNTTGGDN 338
Db 391 SGAVSGTRSDGKAFTSISGG--QADALMLEKSGSFTLNAGDTATDTTVNGGLFTARGTL 448

QY 339 VGVVNL-SGALSQVGTGNIGNTNSLATINIG-----AGVATLDGAVIKATTTKLTDDASVL 392
Db 449 AGTTTLNNGAILTLSGKTVNNDTL-TIREGDALLQGGSLTNGSVKSGSOTLT----- 501

QY 393 IFTNPVVVTGAIDNTGNANKGVVFTGASTVTDNIGNTAVLAESVCGALLOIQGGVYKA 452
Db 502 -----VSNITLTQKAVNLNEGTLTNDSTVTDVTDIAQGTGTA--LKLGTSTVLN 547

QY 453 NAI-----NLTDNASVYFTGDSVTGTSI-----GGTELFATVNT----- 487
Db 548 GAIDPTNVTLASGATWNPONATVQSVVDDLSHAGQIHFTSTRTGKVPATLKVKNLNGQ 607

QY 488 GAGITLRLAGSLLANNIDFGAASNLFPNGPKYNYLIGTIANGNAT-LININAAGTVIA 546
Db 608 NGTISLRVPRDPMQANNAD-----RLVIDGGRATGKTTLNLVYNAGNSAGSLATSKGQIV 662

QY 547 .NDVSIGTVQAO-INTQNKI 564
Db 663 EAINGATTEGAFVQGNRL 681

RESULT 15
SLAP_CAMEE
ID SLAP_CAMEE STANDARD; PRT; 933 AA.
AC P35827;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE S-layer protein (Surface array protein) (SAP).

GN SAPA.
OS Campylobacter fetus.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=196;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
RC STRAIN=84-32 / 23D;
RA MEDLINE=90354448; PubMed=2387868;
RX Blaser M.J., Gotschlich E.C.;
RT "Surface array protein of Campylobacter fetus. Cloning and gene
structure.";
RL J. Biol. Chem. 265:14529-14535(1990).
RN [2]
RP ERRATUM.
RX MEDLINE=91035477; PubMed=2229082;
RA Blaser M.J., Gotschlich E.C.;
RL J. Biol. Chem. 265:19372-19372(1990).
CC -|- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. THIS PROTEIN IS
CRITICAL FOR VIRULENCE.
CC -|- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
S-LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
KW EMBL; J05577; AAA23032.1; -.
DR Cell wall; S-layer.
SQ SEQUENCE 933 AA; 96757 MW; F88C729B4BA5B1E9 CRC64;

Query Match 8.3%; Score 250; DB 1; Length 933;
Best Local Similarity 23.2%; Pred. No. 6.6e-05;
Matches 163; Conservative 87; Mismatches 284; Indels 168; Gaps 33;

QY 5 SLKLFQK-----ATQGLIKLTALTFTTAAIMLTGSG-VLGAARTVTADGAEL-----AAGT 54
Db 229 TLSAFDKTRSVGLGRDITLNAIFTAITRAALLTDQAEIITKRTNVENINISLDTSGD 288

QY 55 NIGPG---AGAFVAGSTLQYTGFTVTDADVSVRALDNNFAAGLFSVTG--DISLGSVV 109
Db 289 FVNGYKGVFNVLGDIIVSF-----ATDASKSVNVTGTITTAAGTCKVDVVAGKIS 343

QY 110 DTGANKLAVNI---DDGLTLTLTGCTAA---YGANPALLFOGGOAAANNVTYALGNI 162
Db 344 ALTADSTSVNLATNNDITITLSANAATSVNLKQKQAKDATITTSAMQOKYNNRRNRATI 403

QY 163 TLGGANAGLTIASDPDVLGPITLGNIDGGIITDNTDAAINGTIGNTPAAQISI-GAS 221
Db 404 TSATAVENLVKHAATNVA-----LNGG-----MDKLATVTLDNAALTAADIKSAS 449

QY 222 TSLT-----GGAVIKATTTKLTNNAAPVLTLTNAVAV----- 252
Db 450 TLNLINSVNGPKHLYSSKRYCKFKRAAAKVKLNTTAATDQTVTL---KANATDNSLEF 506

QY 253 --LTGADVNTTGGDDVGVNLNGALSQVGTGNIGNT--NSLATISVG----- 294
Db 507 DSATAKTSVTASGSGTKLVKGAETLVNIDTFAFNALQSVSFGTGGGKFSVKTGT 566

QY 295 -----AGTATLGGAVIKA---TTTKLTNAASVLTLTNAVLTGADVNTTGGDNVGVNVL 344
Db 567 GDDKIEFVGTLLTEGSDVDPAGNDTIAMKSAA--LTSANFTMIKNIEVNAISDAVATDL 624

QY 345 SGALSO-----VTGNIGNT-----NSLATINIGAGVATLDGAVIKATTTKLTDDASVLIFT 395
Db 625 SSSAFKNSVIITTKAAADTTLTINKDQVINFTAA---DAGSVKLTIVKLVNDVLTALMIV 680

QY 396 NPVVVTGAID-----NTGNANKGVVIFTGASTVTDNIGNTAVLAESVCGALLOIQGGV 450

```
Db 681 KIVLDAAKDTNIALGTAADKALVIDTGTETL--NI-----TSLVKATSPET 726
QY 451 KANAIN--LTDNASVVTFTGDSVTGSG--GTELEATVNICAGITLRAGGSLAANNIDF 506
Db 727 TANTVNAKLTDTVTSLI-IDGMQITLGHAGTAGTD-YSKVSMIDASALKAGLTFDASAILL 784
QY 507 GAASNLEFNGPAG-----KNYNLIGTIANGNNATLNIINAAGTVIANDVYSIGTVAQINIQ 560
Db 785 GA--NATIKGGGADSITVKGGINVVVDLYAGGDDTITLKG-----AEKTDITTVNNFN-A 837
QY 561 NNKIFVINAKN-----ADVDIILDAQAISFKGAAS 589
Db 838 GDKIDIADAKNGTFTFNKITMNSDANLDDYITKAVAGDGSN 879
```

Search completed: April 14, 2003, 16:09:26
Job time : 39 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 16:07:09 ; Search time 94 Seconds
(without alignments)
1356.842 Million cell updates/sec

Title: US-09-800-065-2

Perfect score: 3022

Sequence: 1 MANISLKFQKAIQKGLKTA.....MIELSLKIIYPVLLTVVVS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.rodent.*
- 12: sp.virus.*
- 13: sp.unclassified.*
- 14: sp.virus.*
- 15: sp.bacteriap.*
- 16: sp.bacteriap.*
- 17: sp.archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3022	100.0	619	2 Q9F9H4	Q9F9H4 rickettsia
2	1223.5	40.5	834	2 Q52647	Q52647 rickettsia
3	1103	36.5	1029	2 Q52708	Q52708 rickettsia
4	952	31.5	192	2 Q9F0Q0	Q9F0Q0 rickettsia
5	710.5	23.5	2106	2 Q9XC47	Q9XC47 rickettsia
6	452.5	15.0	2340	16 Q9ZD91	Q9ZD91 rickettsia
7	422.5	14.0	1604	2 Q9K999	Q9K999 rickettsia
8	376	12.4	1617	2 Q9K8B9	Q9K8B9 rickettsia
9	372	12.3	210	2 Q9KJH3	Q9KJH3 rickettsia
10	366.5	12.1	1618	2 Q9KKB1	Q9KKB1 rickettsia
11	366	12.1	1654	2 Q93QW9	Q93QW9 rickettsia
12	365	12.1	1616	2 Q9KKB0	Q9KKB0 rickettsia
13	363.5	12.0	1616	2 Q9KKA5	Q9KKA5 rickettsia
14	361.5	12.0	3165	16 Q8XY13	Q8XY13 rickettsia
15	361	11.9	1615	2 Q9KKA8	Q9KKA8 rickettsia
16	360.5	11.9	2691	16 Q8XPUI	Q8XPUI rickettsia

17	359.5	11.9	2751	16 Q8XUK0	Q8XUK0 rickettsia
18	356.5	11.8	2737	16 Q8XPUI	Q8XPUI rickettsia
19	354	11.7	1618	2 Q9KKB4	Q9KKB4 rickettsia
20	354	11.7	3930	16 Q98E20	Q98E20 rickettsia
21	353.5	11.7	667	16 Q8XRM8	Q8XRM8 rickettsia
22	353	11.7	1615	2 Q9F0P9	Q9F0P9 rickettsia
23	338	11.2	3552	16 Q8XSD6	Q8XSD6 rickettsia
24	336	11.1	1643	2 Q9F0P7	Q9F0P7 rickettsia
25	335.5	11.1	204	2 Q9FDP2	Q9FDP2 rickettsia
26	335	11.1	1643	2 Q9F0P6	Q9F0P6 rickettsia
27	331	11.0	3501	16 Q8Y106	Q8Y106 rickettsia
28	330.5	10.9	1616	2 Q9KKB8	Q9KKB8 rickettsia
29	329.5	10.9	197	2 Q56384	Q56384 rickettsia
30	329.5	10.9	2432	12 Q8QZ06	Q8QZ06 chilo iride
31	329.5	10.9	3705	16 Q8ZHA1	Q8ZHA1 versinia pe
32	329	10.9	196	2 Q9LAQ9	Q9LAQ9 rickettsia
33	329	10.9	2554	16 Q9YDM7	Q9YDM7 brucella me
34	328.5	10.9	204	2 Q9F9R4	Q9F9R4 rickettsia
35	328	10.9	196	2 Q9LA08	Q9LA08 rickettsia
36	328	10.9	1620	2 Q9KKB5	Q9KKB5 rickettsia
37	327.5	10.8	1616	2 Q9KKA2	Q9KKA2 rickettsia
38	327.5	10.8	1616	2 Q9KKA1	Q9KKA1 rickettsia
39	327	10.8	196	2 Q9LAR2	Q9LAR2 rickettsia
40	326.5	10.8	197	2 Q57548	Q57548 rickettsia
41	326.5	10.8	1616	2 Q9KKB3	Q9KKB3 rickettsia
42	326	10.8	1616	2 Q9KKA0	Q9KKA0 rickettsia
43	325.5	10.8	3705	2 Q9F285	Q9F285 versinia pe
44	324.5	10.7	197	2 Q52655	Q52655 rickettsia
45	324.5	10.7	204	2 Q9F9R6	Q9F9R6 rickettsia

ALIGNMENTS

RESULT 1

Q9F9H4 PRELIMINARY; PRT; 619 AA.
 ID Q9F9H4;
 AC Q9F9H4;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Outer membrane protein A.
 OS OMPA.
 GN Rickettsia felis (Rickettsia azadi).
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales.
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_taxid=42862;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21217364; PubMed=11321078;
 RA Bouyer D.H., Stenos J., Crocquet-Valdes P., Moron C.G., Popov V.L.,
 RA Zavala-Velazquez J.E., Foil L.D., Stothard D.R., Azad A.F.,
 RA Walker D.H.;
 RT "Rickettsia felis: molecular characterization of a new member of the
 RT spotted fever group.";
 RL Int. J. Syst. Evol. Microbiol. 51:339-347(2001).
 DR EMBL: AF191026; AAG17185.1;
 SQ SEQUENCE 619 AA; 60547 MW; 1AA112066EEFA1B CRC64;

Query Match 100.0%; Score 3022; DB 2; Length 619;
 Best Local Similarity 100.0%; Pred. No. 8.7e-110;
 Matches 619; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MANISLKFQKAIQKGLKTALETTSTAATMLTGSVGLGAARTVTADGAELAAGTNGPCA 60
 DB 1 MANISLKFQKAIQKGLKTALETTSTAATMLTGSVGLGAARTVTADGAELAAGTNGPCA 60
 QY 61 GAFVAGSTLQYTGAFVTVDADVSVALDLNNEFAAGLFSVTGDISLGSVYDTGGANKLAVN 120
 DB 61 GAFVAGSTLQYTGAFVTVDADVSVALDLNNEFAAGLFSVTGDISLGSVYDTGGANKLAVN 120
 QY 121 IDDGTLTLTGTGTAAAYGANPALFFGGQAAANNNTYALGNITLGGANAGLTIASDPDVL 180
 IDDGTLTLTGTGTAAAYGANPALFFGGQAAANNNTYALGNITLGGANAGLTIASDPDVL 180

Db 121 IDGLTULTLTGTAAYGANPALLFQGGAAANNYYTALGNITLGGANAGLTIASDPVL 180
 Qy 181 GPITLAGNIDGGIITDNTDAINGTICNTNPAQAISICASTLSLGGAVIKATTTKLTNA 240
 Db 181 GPITLAGNIDGGIITDNTDAINGTICNTNPAQAISICASTLSLGGAVIKATTTKLTNA 240
 Qy 241 APVLTLTNANAVLTGAVDNTTGGDDVGVNLNGALSQVGTGNIGNTNSLATISVGAGTATL 300
 Db 241 APVLTLTNANAVLTGAVDNTTGGDDVGVNLNGALSQVGTGNIGNTNSLATISVGAGTATL 300
 Qy 301 GGAVIKATTTKLTNAASVLTLTNANAVLTGAVDNTTGGDNGVNVNLSGALSQVGTGNIGNTNS 360
 Db 301 GGAVIKATTTKLTNAASVLTLTNANAVLTGAVDNTTGGDNGVNVNLSGALSQVGTGNIGNTNS 360
 Qy 361 LATINIGAGVATLGGAVIKATTTKLTDDASVLIETNPVVVTCADTGNANKGVVIFTGA 420
 Db 361 LATINIGAGVATLGGAVIKATTTKLTDDASVLIETNPVVVTCADTGNANKGVVIFTGA 420
 Qy 421 STVTDNTGNITAVLAESVVGALLQIQGGVVKANAINLTDNASVVTFTGDSVTGSGIGTE 480
 Db 421 STVTDNTGNITAVLAESVVGALLQIQGGVVKANAINLTDNASVVTFTGDSVTGSGIGTE 480
 Qy 481 LFAVTVNIGAGTTLRAGGSLAANNIDFGAASNLFEFNGPAGKKNYLTGTIANGNNATLINA 540
 Db 481 LFAVTVNIGAGTTLRAGGSLAANNIDFGAASNLFEFNGPAGKKNYLTGTIANGNNATLINA 540
 Qy 541 AGTVIANDVSGTVAQINIQNKIFVINAKNADVDIILDAQAISFKGAASRLFLANVSLOM 600
 Db 541 AGTVIANDVSGTVAQINIQNKIFVINAKNADVDIILDAQAISFKGAASRLFLANVSLOM 600
 Qy 601 IELSLLKIIYPVLTVVVS 619
 Db 601 IELSLLKIIYPVLTVVVS 619

RESULT 2
 Q52647 PRELIMINARY; PRT; 834 AA.
 AC Q52647
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Outer membrane protein A (Fragment).
 GN OMPA.
 OS Rickettsia akari.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=786;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KAPLAN;
 RX MEDLINE=93194085; PubMed=7680636;
 RA Gilmore R.D.Jr. ;
 RT "Comparison of the ompA gene repeat regions of Rickettsiae reveals
 species-specific arrangements of individual repeating units.";
 RL Gene 125:97-102(1993)
 DR EMBL; L01461; AAA26387.1; --
 FT NON_TER 834 834
 SQ SEQUENCE 834 AA; 79350 MW; 703F3BAE9B51176C CRC64;

Query Match 40.5%; Score 1223.5; DB 2; Length 834;
 Best Local Similarity 51.3%; Pred. No. 3.5e-40;
 Matches 307; Conservative 46; Mismatches 93; Indels 153; Gaps 19;

Qy 20 ALFTTSTAAILTGS-----GVL---GAARTVTAD--GAELAAGTNIIGPG-----A 60
 Db 261 SVLTLTNANAVLTGAIDNTTGVNDVGVNLNGALSQVGTGNIGNTNSLATISVGAGTATL 320
 Qy 61 GAFVAGTLOQTGA---FTVTDADY-----SVRALDNLNFAAGLFSVTGDIS 104
 Db 321 GAVIKATTTDLTNAASVLTLTNANAVLTGAIDNTTGVNDVGVNLN---GALSQVTGNI- 376

Qy 105 LGSVVDTGANKLAVNIDDDGLTLTLTGTAAYGANPALLFQGGAAANNYYTALGNITL 164
 Db 377 -----GNTNSLA-----TISV-GAGTATLGG--AVI---KATTTDLTDAASVLT- 415
 Qy 165 GGANAGLTIASD---PDVLPITL-----AGNID-----GGGII- 195
 Db 416 TNANAVLTGAIDNTTGVNDVGVNLNGALSQVGTGNIGNTNSLATISVGAGTATLGGAVIK 475
 Qy 196 ---TDNTDAA-----INGTIGNTPAAQI 216
 Db 476 ATTTDLTDAASVLTLTNANAVLTGAIDNTTGVNDVGVNLNGALSQVGTGNIGNTNSLATI 535
 Qy 217 SIGASTLSLGGAVIKATTTKLTNAAPVLTLTNANAVLTGAVDNTTGGDDVGVNLNGALS 276
 Db 536 SVGAGTATLGGAVIKATTTDLTDAASVLTLTNANAVLTGAIDNTTGVNDVGVNLNGALS 595
 Qy 277 QVTGNIGNTNSLATISVGAGTATLGGAVIKATTTKLTNAASVLTLT--NAVLTGAVDNTT 334
 Db 596 QVTGNIGNTNSLATISVGAGTATLGGAVIKATTTDLTDAASVLTLTNANAVLTGVIDNTT 655
 Qy 335 GGDVNVVNLGALSQVGTGNIGNTNSLATINIGAGVATLGDGAVIKATTTKLTDDASVLIF 394
 Db 656 GVDNVGVNLNGALSQVGTGNIGNTNSLATISVGAGTATLGGAVIKATTTKLTDDASQVTF 715
 Qy 395 TNPVVVTGAIDNTGNANKGVVIFTGASTVDNIGNTAVIAEVSVCAGLLQIQGGVVKANA 454
 Db 716 TNPVVVTGAIDNTGNANNGIATFTGNSTVGTGNIGNTAALATVNVGAGLLQVGGVVKANA 775
 Qy 455 INLTDNASV-----TFTGDSVTGSGIGTELEFATVNIAG 490
 Db 776 INLTDNVSAVTTPNPVVVTGAIDNTGNANNGIATFTGNSTVGTGNIGNTAALATVNVGAG 834

RESULT 3
 Q52708 PRELIMINARY; PRT; 1029 AA.
 AC Q52708
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Outer membrane protein A (Fragment).
 GN OMPA.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KENYA TICK TYPHUS;
 RX MEDLINE=93194085; PubMed=7680636;
 RA Gilmore R.D. Jr. ;
 RT "Comparison of the ompA gene repeat regions of Rickettsiae reveals
 species-specific arrangements of individual repeating units.";
 RL Gene 125:97-102(1993)
 DR EMBL; L01462; AAA99908.1; --
 FT NON_TER 1029 1029
 SQ SEQUENCE 1029 AA; 99999 MW; 09573881A5B9BDC1 CRC64;

Query Match 36.5%; Score 1103; DB 2; Length 1029;
 Best Local Similarity 42.3%; Pred. No. 1.9e-35;
 Matches 283; Conservative 60; Mismatches 194; Indels 132; Gaps 18;

Qy 16 GLKTALETTSTAAILTLT--GSGVLGAARTVTADGAELAAGTNIIGPGAF-----VAGSTL 69
 Db 365 GNTNALATISVGAGKATLGGAIKATTTKLTDNASQVTFTPNPVVVTGAIDNTGNANNGIV 424
 Qy 70 QYGAFTVTADVSVRALDNLNFAAGLFSVTGD-----ISLGS 107
 Db 425 TTTGGSTVGTGNIGNTALATVNVGAGLLRVQGGVVKNSNTINLTDNASQVTFTPNPVVVTGA 484
 Qy 108 VDTGGANKLAVNIDDDGLTLT-----LTGTGTAAG----- 138


```

Db 485 IDNTGNANGIVTFTGDSVTGNTGNATATISVGAGKATLGGAIKATTTKLTNDASQ 544
QY 139 ---ANPALL-----FQGOAANN--TYTALGNITLGGANAGL----- 171
Db 545 VTFTNPVVVGTGADNTGNANGIVTFTGSGVTGNTGNATATYVVG---AGLLRVGG 601
QY 172 -----TI-----ASDPDLVGPITLAGNID-----GGGIITDNTDAAINGTINPAAQI 216
Db 602 VYKSYNTINLTNDASQVTFNPVVVGTGADNTGNANGIVTFTGDSVTGNTGNATATV 661
QY 217 STGASTLSGGAVIKATTTKLTNAAPVLTNANAVLTGAVDNTTGGDDVGVNLNGALS 276
Db 662 NVGAGLLRVGGVWKSNTINLTNDASQVTFNP--VVVTGAIDN--TGNANGIVTFTGN--S 718
QY 277 QYTGNTGNATATISVGAGTATLGGAVIKATTTKLTNAASVLTLT--NAVLTGAVDNTT 334
Db 719 TVTGNIGNTALATVNVGAGIATLGGAVIKATTTKLTNAASVLTLTNNAVLTGAIDNT 778
QY 335 GGDNVGVNLGALSQVGTGNTGNATATINIGAGVATLDGAVIKATTTKLTDDASVLIF 394
Db 779 GVDNVGVNLGALSQVGTGNTGNATATISVGAGKATLGGAVIKATTTKLTNDASQVTF 838
QY 395 TNPVVVGTGADNTGNANGVITFTGASTVTDNIGNTAVLAESVSGAGLIQIGGVYKANA 454
Db 839 TNPVVVGTGADNTGNANGIATFTGDSVTGNTGNATATVNVGAGLLRVGGVWKSNT 898
QY 455 INLTNDAS-----VVTFGDSVTGSGIGCTELFATVNIAGI 491
Db 899 INLTNDASAVTNPVVVGTGADNTGNANGIVTFTGDSVTGNTGNATATISVGAGK 958
QY 492 TLRAGSGLAANNIDF--GAASNLEFNGPAGKNYLNIGTIANGNNATININAAGTVIANDVS 550
Db 959 ATLGGAIKATTTKLTNDASQVTFNPVVV--CAIDNTGNANGIVTFTGDSVTGNTGN 1017
QY 551 IGTVAQINI 559
Db 1018 TNALATVNV 1026

RESULT 4
Q9F0Q0 PRELIMINARY; PRT; 192 AA.
AC Q9F0Q0;
DT 01-MAR-2001 (TREMREL..16, Created)
DT 01-MAR-2001 (TREMREL..16, Last sequence update)
DE Outer membrane protein OmpA (Fragment).
GN OMPA.
OS Rickettsia sp. California 2.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=147259;
RN [1]
RP SEQUENCE FROM N.A.
RA Roult D.;
RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-CALIFORNIA 2;
RA Roult D.;
RL "A new SFG rickettsia isolated from fleas.";
RN Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-CALIFORNIA 2;
RA Roult D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF210694; AAG48555.1;
FT NON_TER 1
FT SEQUENCE 192 AA; 18596 MW; A9E1C76C41D4590D CRC64;

Query Match 31.5%; Score 952; DB 2; Length 192;
Best Local Similarity 100.0%; Pred. No. 2.6e-30;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LFQKAIQKGLKLTALTSTTAAIMLTGSGVLGAARTVTADGAELAAAGTNIQPGAGAFVAGS 67

```

```

Db 1 LFQKAIQKGLKLTALTSTTAAIMLTGSGVLGAARTVTADGAELAAAGTNIQPGAGAFVAGS 60
QY 68 TLQYTGAFVTVDADSVSRALDNLNFAAGLSVGTGDISLGSVVDTGANKLAVNIDGLT 127
Db 61 TLQYTGAFVTVDADSVSRALDNLNFAAGLSVGTGDISLGSVVDTGANKLAVNIDGLT 120
QY 128 TLTGCTAAYGANPALLFQGGQAAANNYYTALGNTTLGGANAGLTATASDPDLVGPITLAG 187
Db 121 TLTGCTAAYGANPALLFQGGQAAANNYYTALGNTTLGGANAGLTATASDPDLVGPITLAG 180
QY 188 NIDGGGIITDNT 199
Db 181 NIDGGGIITDNT 192

RESULT 5
Q9XC47 PRELIMINARY; PRT; 2106 AA.
AC Q9XC47;
DT 01-NOV-1999 (TREMREL..12, Created)
DT 01-MAR-2001 (TREMREL..16, Last sequence update)
DT 01-DEC-2001 (TREMREL..19, Last annotation update)
DE Outer membrane protein A.
GN OMPA.
OS Rickettsia australis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=787;
RN [1]
RP SEQUENCE FROM N.A.
RA Stenos J.; Walker D.H.;
RC STRAIN-PHS;
RC MEDLINE=20487299; PubMed=11034486;
RT "The rickettsial outer-membrane protein A and B genes of Rickettsia australis, the most divergent rickettsia of the spotted fever group.";
RL Int. J. Syst. Evol. Microbiol. 50:1775-1779(2000).
DR EMBL; AF149108; AAD39531.2;
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
SQ SEQUENCE 2106 AA; 209075 MW; A12B28F54BEEDECC CRC64;

Query Match 23.5%; Score 710.5; DB 2; Length 2106;
Best Local Similarity 25.5%; Pred. No. 5.6e-20;
Matches 300; Conservative 79; Mismatches 217; Indels 579; Gaps 42;

QY 1 MANISLKLFQKAIQKGLKLTALTSTTAAIMLTGSGVLGA-----RTVTADG 47
Db 1 MANISPKLFQKAIHKSLKALFTTSTAAIMLSSSGAWGAAGVISVNDAAFSNRVANW 60
QY 48 AELAACTNIGPGAGAFVAG-----STLQYTGAFVTVDADSVSRALDNLNFAAG---LFSVT 100
Db 61 NEITAG---GAANGNHADGPQDNEAFYGGNHTIT--ADEAGRIITAINVAGITPVALNST 116
QY 101 GDISLGSVVDTGANKLAVNIDGLTTLTGTTAAYGANPALLFQGGQAAANNYYTALG 160
Db 117 QNTSVGSIV-TGG-NLLPVTIADGKSLTTLTGTKAVALD-----HGFDAAADN-YTGLG 166
QY 161 NITLGGANAGLTATASDPDLVGPITLAGNIDGG-----GIITDNTDAINGTIGNT----- 210
Db 167 AIALGGATAGLTQSATPAL--INLAGAIDGNNNDHGEITVNTRTSFTGVIGTKTLHAV 224
QY 211 -----NPAAGISIG----- 219
Db 225 TFNNGNAGVSTAGAISATVVMIGEDAGNVASVVQTLGITGAVNFAADGALTANNIAGA 284
QY 220 -----ASTLSLGG-----AVIKATTTKLTNAAPVLT 246
Db 285 VTANNTGTLTVGAGDVTGAICVNGNLIKQVLFNGASNVATIDATNTVINNAANVTA 344
QY 247 T-----NANAVLTGAV---DNTTG-----GDDVGVNLNGA--LSQV-- 278
Db 345 AGAITAANFAADGALTANNIAGAVTTANNNTGTLTVGAGDVTGAICVNGNLIKQVLF 404

```

```

QY 279 -----TCNIG-----NTNSLAT 290
Db 405 NGASNVATIDATNVTINNAANVTAGAITAAVFAADGALTANNGIAGAVTTANNCT 464
QY 291 ISVGATAT-----LGGAVIK-----ATTIKLNAASVLTLTNAV-----325
Db 465 LTVGAGDVTGALGVNGNLLKQVLFNGASNVATIDATTVTINNVAANVTAGAITAAVNF 524
QY 326 -----LTVGAV-----DNTTGGDNVGVNLSGALSOVTGNI-----355
Db 525 AADGALTANNGIAGAVTTANNCTGTLTVGAGDVTGALGVNGNLLKQVLFNGASNVATID 584
QY 356 -----GNPNSLATINIGAVAT-----372
Db 585 ATNVTINNAANVTAGAITAAVFAADGALTANNGIAGAVTTANNCTGTLTVGAGDVTG 644
QY 373 ---LDG-----AVIKATTKLTDDAS-----VLIFTN 396
Db 645 AIGVNGNLLKQVLFNGASNVATIDATTVTINNVAANVTAGAITAAVFAADGALTANN 704
QY 397 PVV-----VTGAID-NTGN-----409
Db 705 GIVGAVTTANNCTGTLTVGAGDVTGALGVNGNLLKQVLFNGASNVATIDATNVTINNA 764
QY 410 -----ANKGV-----414
Db 765 ANVTAGAITAAVFAADGALTANNGIAGAVTTANNCTGTLTVGAGDVTGALGVNGNLL 824
QY 415 ---VIPTGASTV-----TDNIG-----428
Db 825 KQVLFNGASNVATIDATNVTINNAANVTAGAITAAVFAADGALTANNGIAGAVTTAN 884
QY 429 -NTAVLAESVGA-----GLLOIQGG-----VVKANALNTDNASVVTFTG 468
Db 885 NNTGTL-----TVGAGDVTGALGVNGNLLKQVLFNGASNVATIDATTVTINNVAANVTAA 941
QY 469 DST-----VTGSIG-----477
Db 942 AITAAVFAADGALTANNGIAGAVTTAGAGNGTTLTAGVGTGVTGATNNAASKVLNAKVDN 1001
QY 478 -----GTELEFA-TVNI-----GAGITLRAGGSAAANNIDFGA-----ASNLEFNGPAGK 520
Db 1002 GAADLVFTSDIYARTVNFDFDIAGGAGVGTIQVGGNLIATNVFNGGAGGTLXLNGPVGK 1061
QY 521 NYNLIGTIANGNATLAINAGTIVANDVSTGTVQAOINQONKLFVINAKNADVDILDAQ 580
Db 1062 SYTLSGTIANGNATLAINLGLTVBNAIXIGTVQAOINQONKTFAINVKNADEILNQA 1121
QY 581 AISPKGAASRLFLANVLSQMIELSLKLIYPVLIT 615
Db 1122 AIDFKGANSKFLVNSATDTRVVTIKNDLPAPAT 1156

```

RESULT 6

```

Q92D91 ID Q92D91 PRELIMINARY; PRT; 2340 AA.
AC AC Q92D91;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Cell surface antigen (SCA3).
GN RP451.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Slicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,

```

```

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RL mitochondria"; Nature 396:133-140(1998).
DR EMBL; AJ235271; CAA14908.1;
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Complete proteome.
SQ SEQUENCE 2340 AA; 247862 MW; CC4070F93C165F26 CRC64;

```

```

Query Match 15.0%; Score 452.5; DB 16; Length 2340;
Best Local Similarity 28.5%; Pred. No. 5.7e-10;
Matches 194; Conservative 84; Mismatches 232; Indels 171; Gaps 34;

```

```

...QY 26 TAAIMLTGSGVLGAARTVTADGAELAACTNIGPGAGAFVAGSTLOYTGAFVTD---ADV 82
Db 407 TDCVFNNVNGGTLIINAKNTISAKLLNATKAIQINANLTMNHPAGDISDIRIADN 466
QY 83 SVRALDLNFAAGLFSVTGDISL---GSVVDTGANKLAVNIDDLGLTLTLTGTTAAAYA 139
Db 467 TIYIDAKN-----GNVLLNNAKIIFEGADSMALINTGVADRTFT-----510
QY 140 NPALLFOGQQAANNYYTALGNITLGGANAGLTIASDDVGLGPITLAGN-----ID 190
Db 511 ---IYNLNOSGNDY---GIVKIEAIKVVITANOS---GPYITIGDNTHRLKELIVE 560
QY 191 GGG-IITDNTDAANGTIGTNTNPAQISIGASTLSLGA---VIKATTTKLTNAAPVLT 246
Db 561 GAGDIIDDT---IFTKLLSINSTGQITFN-RTLDLGGAGNIAFGKHTLVNG-----610
QY 247 TNANAVITGAVDNTGGDDVGLNMGALSQVTCNIG-NTNSLATISVAGTATLGGAVI 305
Db 611 -----VTGST--TSENNOGILINS--GNITGVIGTNELGLKLVNIGADPVTCANVF 660
QY 306 KATTTKLTNAASVLTLTNAV-LTGAVDNTGGDNVGVNLSGALSQVTCNIG-NTNSLAT 363
Db 661 --ASVALTNPSVLLIADGVTLTGEV--THNNTKGVLSL-GTGSNITGQITGTSAALEK 715
QY 364 INIGAVATLDGAVIKATTTKLTDDASVLFTNPVVVTGAIDNTGNANKGVVIFTGASTV 423
Db 716 INIGAGASNID-SNIYAGSTVLTDTSELTLLNDVVVNSNIITTAGNSCKLIFTNGGI 774
QY 424 TDNIG-NTAVLAEV-----SVGA-----GLLOI 445
Db 775 TGNIGGAALQGVNFVFGTNTNIGGTANSQNTVAHSAANVITGLTGLKLYKDTGTIIA 834
QY 446 QGVV-----KANALNTDNASV-----VFTFGDSTVTGSIG--GTELEFA 483
Db 835 HGLVGDIDFNKAGKILGDGAMIDSVLCNGVAGTILDFIGDNVTQNIAGDNANSIS 894
QY 484 TVNIGA-----GITLRAGGSAAANNIDFGA-ASNLEFNGPAGKN 521
Db 895 TINIQDNTKNVTIANDIFVDNIHFTNGTILQGLNLTTHNIDFGANGGTLEFNG--NNT 952
QY 522 YNLIGTIANGNATLAINAAGTIVANDVSTGTVQAOIN-----QNNKIFVINAKNADVD 575
Db 953 YNLAIIVNQNGIL--NAETNLKASDDTIGTVKLIINIGOIGTPQN---FTIQVNNKLT 1007
QY 576 ILDA--QAI SFKGAASRLFLA 594
Db 1008 LVSSVNSSINFGDANSOLILS 1028

```

RESULT 7

```

Q9KK99 ID Q9KK99 PRELIMINARY; PRT; 1604 AA.
AC AC Q9KK99;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE OmpB (Fragment).
GN OMPB.
OS Rickettsia helvetica.

```

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OC NCBI_TaxID=35789;

RP SEQUENCE FROM N.A.
 RC STRAIN=C9P9;
 RX MEDLINE=20393643; PubMed=10939649;

RA Roux V., Raoult D.;
 RA MEDLINE=20393643; PubMed=10939649;

RT "Phylogenetic analysis of members of the genus Rickettsia using the
 gene coding the outer-membrane protein ompB (ompB).";
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).

DR EMBL: AF123725; AAF34128.1; -;
 DR InterPro: IPR003858; rOmpA_rOmpB.
 DR Pfam: PF02708; rOmpA_rOmpB; 1.

FT NON_TER 1604 1604
 SQ SEQUENCE 1604 AA; 162714 MW; 83F9AE7597837B32 CRC64;

Query Match 14.0%; Score 422.5; DB 2; Length 1604;
 Best Local Similarity 29.7%; Pred. No. 5.7e-09;
 Matches 194; Conservative 83; Mismatches 230; Indels 147; Gaps 38;

38;

QY 37 LGAARTVTADGAELAACTNIG-PCAGAFVAGSTLOYT---GAFTVT-----DADSV 84

Db 1 MGAA-----NRDINANVLDVIGIVTGDSTLYVVGAGPCTATATCGNDNGGNATV 54

QY 85 RALDLNFAAGLSVTG-DISLGSVDTGGANKLAVNIDDLTLTLTGT-GTAAYGAN-- 140

Db 55 LFGSINLQNGVEAVNGADIAIGSVGTAG-QLLTVNI-AGNTLTNGAPVAFPNVTY 112

QY 141 -----PALLFOGQAA--ANNYTYAL-CN-TLGGANAGLTITASDPVLGPITLA 186

Db 113 TNLGPNVFDAAAFKVSLLAANGADGKTATFNGTATFNGTAAGKGI----- 160

QY 187 GNIDGGIITDNTDAATNGTIGNTPA-----AQISIGASTL---SLGGAVIKATT 235

Db 161 -NIDNGNI-----AFNGTIGNNGIOGLTGTGNAQATLNANKFDTGAGSVSLDNG 213

QY 236 KLTNAAPVLTNANAVLTGAVDNTTGGDDVGLNLNGALSQVGTGNTGNTSLATISVGA 295

Db 214 SILNVADGVNITGAVTNISIDGKAAND--GTVNFGLD-SAVSTDIGANNAATAVNV-A 269

QY 296 GTATL-----GGAVIKATTKLTNAASVLTTL--NAVLTGAVDNTTGGDVGVLNLSGA 347

Db 270 GELTFQGGNNGGGTGTAAINTSAGSLIKFTQDHAVTGNLTNTSGTNNQGGFVNGG 329

QY 348 LSQVGTGNTG-NTNSLATINIGAGVATLDGAVIKATTKLTDDASVLFTNPVVVTGAI-- 404

Db 330 DVTVTGNIGAGNSLATINFE-----TDD-SLIVHKAATNATNALFI 370

QY 405 --DNTGNANKGVVIFTGAS--TVTDNIG-NTAV-----LAESVGA--GLLIQIOGGVVYKANA 454

Db 371 QNYNTATANTGILKLGHTGYAINGNIGANNAALKLVLDLDDNGAATNFTLKQSSIIKAQN 430

QY 455 INLTD-NASVVTFTGDSVTGSL-----GGPELF--ATVNIGAGITLRAGGSL--AA 501

Db 431 ISLADQDNLTLEEGTTITGDIINTKNGNGTIVLTGNATNGGTVGVDIAAGGNLGTI 490

QY 502 NNIDFGAASNLFPNGPAGK-----NYNLIGTIANG-----NNATLINAACTVIANDVSI 553

Db 491 KNIILG-KRDLVSGGNTVQVAPNGINFAANETLSLNTTDPILLASDIITTDKQ-GI 548

QY 554 VAQINTONKIFVINAKNADVILDQAATSFKAASRLFLANVLSQMLIELSLK 607

Db 549 IDGSALTNDQTLTIAGKIGVDIINNPA-----QNVALQQLKIGSSK 590

RESULT 8

Q9KKB9

ID Q9KKB9

AC Q9KKB9

DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE ompB (Fragment).

GN ompB.

OS Rickettsia aeschlimannii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OC NCBI_TaxID=45262;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=MC16;

RX MEDLINE=20393643; PubMed=10939649;

RA Roux V., Raoult D.;

RT "Phylogenetic analysis of members of the genus Rickettsia using the

gene coding the outer-membrane protein ompB (ompB).";

RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).

DR EMBL: AF123705; AAF34108.1; -;

DR InterPro: IPR003858; rOmpA_rOmpB.

DR Pfam: PF02708; rOmpA_rOmpB; 1.

FT NON_TER 1617 1617

SQ SEQUENCE 1617 AA; 163501 MW; B7A1B5DB1403BE62 CRC64;

Query Match 12.4%; Score 376; DB 2; Length 1617;

Best Local Similarity 25.0%; Pred. No. 3.6e-07;

Matches 168; Conservative 80; Mismatches 241; Indels 184; Gaps 33;

QY 38 GAARTVTADGA-----ELAAGTNIGPGAGAFVAGSTLOYTGAFVTDDADSVRALDL--NN 91

Db 13 GAATTV-DGAGFDQTAAPANVAVALNAVI---TANANGINLNTPDGSEFNGFLDTANN 67

QY 92 FAAGLSVTGDISLGSVD--TGGANKLAVNIDDLTLTLTGTG-TAAYGA-----NPAL 143

Db 68 LAV--TVSADTTLGFTITNAANNANSFDLRVDAGKTLTITGGQITTAQAAVTKNAQNVVA 124

QY 144 LFOGQAAANNYYTALGNITLGGANAGL--TIASDDPVLGPTPLAGNIDGGGIITDNTDA 201

Db 125 QFNGGAAIANNDLSGVGTIDLGAASATLVFNLANPTTKAPLVLGN--AVIAN-- 176

QY 202 AINGTIGTNTNPAQAIS-----IG-----ASTLSLGGAVIKATT 235

Db 177 GYNGTLNVINGFLOVSDKSFATPKVTINIGDQGFIENTDATVGNALNLQAGGATINFGT 236

QY 236 KLTNAAPVLTNANAVLTGAVDNTTGGDDVGLVNLN--GALSQVGTGNTGNTSLATISV 293

Db 237 DGTGR--LVLLSKHGAANDFNVTGSLGGLKGVIAFNTVAVAGQLIANAGPANAVIGTNN 294

QY 294 GAG-----TATLGG-----AVIKAT-----TTKLTNAA 316

Db 295 GAGRAAGFVVSDNGNAATIAGOVYAKOMVIOANAGGEVNFHIVDVGTDGTAFKTA 354

QY 317 SVLTLTNAVLTGAVD-----NTTG-----GDNVGVVNLSG-----ALS 349

Db 355 SKVAITONSFGTDFGNLAAQITVPTDTWLTGNTGTDSNHNHTAGVITFAANGTLASA 414

QY 350 QVTGNTGNTSLATINI--GAGVATLDGAVIKATTKLTDDASVLIFTNPVVVTGAI DNT- 407

Db 415 SADANVAVTNNITAEASGIVGVQLSGT--HTAELRLGNAGSIFKLADGTGVINGKVNQA 472

QY 408 ---GNANKGVVIFTGASTVTDNIGTAVLAESVSGVGLLIQIOGGVVYKANAINTDNASV 464

Db 473 VYGGALAAAGATLDGSAITGDI--GGVAALOG-----I 507

QY 465 TFGDSTVTSIGGTGELFATVNIAGITLRA--GGSLAANN-----IDFGAASNLFPNGP 517

Db 508 TLANDATKTLTGGANIIG--ANVGGTIDQANGGKIKLTNTONNIIIVDFDLAIATDQTV 566

QY 518 A-----GKNYNLTGTTA--NGNATL--NINAA-----GTVIANDSVSTGVAQINI 559

Db 567 VDASLTNAQTTLTISGTGAIGANNKTLQGFNIGSKTALNGMVAINELVIGNSVQF 626

QY 560 QNNKIFVINAKNA 572

Db 627 AHNTYLTSTTNA 639

RESULT 9

Q9KJH3 PRELIMINARY; PRT; 210 AA.
AC Q9KJH3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Outer membrane protein A (Fragment).
OS Rickettsia sp. FUJ98.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=116094;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FUJ98;
RA Zhang P.H., Cao W.C., Zhang X.T., Xu R.M., Dai X.H., Gao D.Q.;
RT "Identification of Spotted Fever Group Rickettsiae in ticks from
RT Southern China".
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF169629; AAF86358.1;
FT NON_TER 210
SQ SEQUENCE 210 AA; 5BA5DA67DAAC82B CRC64;

Query Match 12.38; Score 372; DB 2; Length 210;
Best Local Similarity 50.28; Pred. No. 7.1e-08;
Matches 113; Conservative 15; Mismatches 63; Indels 34; Gaps 11;
Qy 1 MANISLKLFOKAIQGLKLTFTSTAAIMLTGSGVLGAARTVTA-----DGAEE--- 49
Db 1 MANISPKLFQKAIQGLKLTFTSTAAIMLTSSGVLGAAGVATNDAAAFSNDAAEANN 60
Qy 50 ----LAAGTNGPGAGAFVAGSTTQYTGAFVTDADVSVRALDNNFA-----AGLFSVTG 101
Db 61 WDEITAEGVANGIPAGGQNNWAFYSADYIT-ADVVDRIITAINVAGTTPVGL-NIAQ 118
Qy 102 DISLGSVVDTCGANKLAVNIDGLTTLTGTAAGANPALLFOGQAAANNVYTLGN 161
Db 119 NTVGSIITRG--NLLPVTI-AGKSLTLNGTNAVA--AN-----HGPDPADN-YTGLGN 167
Qy 162 ITLGGANAGLTASDPVGLGPITLAGNIDGGIITNDTDAINGT 206
Db 168 ITLGGANAELIQSATP--AKITLAGNIDGGIITVTKDAAINGT 210

RESULT 10

Q9KKB1 PRELIMINARY; PRT; 1618 AA.
AC Q9KKB1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE OmpB (Fragment).
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RT gene coding the outer-membrane protein ompB (ompB)".
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455 (2000).
DR EMBL; AF123713; AAF34116.1;
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
FT NON_TER 1618
SQ SEQUENCE 1618 AA; 164193 MW; 66AD2199620750C6 CRC64;
Query Match 12.18; Score 366.5; DB 2; Length 1618;
Best Local Similarity 24.18; Pred. No. 8.4e-07;

Matches 179; Conservative 91; Mismatches 251; Indels 223; Gaps 37;
Qy 23 TTSTAAILMTGSG---VLGAARTVTDAGELAA-----GTNIGPGAGAFVAGSTLQYTGAF 75
Db 10 TTNGVATTVDSGVGFDQTVLANVAVAPNAVITANANNINGINLNTPAGSF-----NGLF 61
Qy 76 TVTDADVSRALDLNPAAGLFSVTGDISLGSVVD-TGGANKLAVNIDDLGLTLTLTG- 133
Db 62 -----LSNANLAV---TVSEDTTLGFINNAANNANRNLTLDAKTLITITGGI 108
Qy 134 ----TAAAGA-PPALLFOGQAAANNVYALGNITLGGANAGLT---IASDPVVLGPITL 185
Db 109 TNVOSAATHNAQNIIVAKENGAAIANDLSGLGTIDFGAAASTLVFLANPTQKAPLIL 168
Qy 186 AGNIDGGIITDNTDAAINGTIGNTPAAQIS-----IG----- 219
Db 169 ADN---ALIVNGA---NGTLNVNTNGFIQVSDKSPATVKAINIGDGGQGFMTNATNAN 220
Qy 220 ASTLSLGGAVIKATTTKLTAAPVLTLTNANAVLTG-ADVNTTGGDDVGVNLNGLAL--S 276
Db 221 ALNLQAGGTTINFNGTDGTGR---LVLLSKNGAATDFNVTGSLGNNLKGIIELNTVAING 277
Qy 277 QVTNIGNTNSLATISVGAG-----TATLGG-----AVIKAT----- 308
Db 278 QLINAGPANAVICTNNGAGRAAGFVVDNGKRAATIDGQVYAKDMVIOSSANANGQVNER 337
Qy 309 -----TTKLTNAASVLTLTNAVLTGAVD-----NTTG-----G 336
Db 338 HIVDVGIDGTTAFKTAASIVAITQNSNFGTTDFGNLAAQVTPDMLTGNFTGDANNPG 397
Qy 337 DNVGVNLSG---ALSOVTCNIGNTNSLATINI-GAGVATLDGAVIKATTTKLTDDASV 391
Db 398 NTAGVITFAANGTLASASADANNAVITNITAEASGVGVVQLSGT---HTAELRLGNAGSV 455
Qy 392 LIFFNVVVVTAIDNT---GNANKGVVIETGASTVTDNIGN---TAVLAESV----- 438
Db 456 FKLADGVINGKVNQTVLVGVLAAGAITLDGSAITITDIGNGGGGAALOSITLANDATK 515
Qy 439 -----GAGLIQIQGVVKANA-----INLTD-----NASVVTFTG 468
Db 516 TLTLLGGAIIISANGTTFNFQANGTKIKLTQNNIVVDCDLAIATDQTVVDASSLTNAQ 575
Qy 469 DSTVTGSGIS---GTFLFATVNIAGAITLRAGGSLAANNIDFGAASNLEF----- 514
Db 576 TLTISGTIGIIGANTTLGQFNIGSSKTYTLNGGNVAINELVIGNNGSVQFAHNTLYLRT 635
Qy 515 NGPAGKNYNLIGTIANGNNATLINAAGTVIANDVSTGTVQAQINQNNKIFVINAKNADV 574
Db 636 TNAAGQCKIIFNPVNV-NNTTL---AAGTNLGS---AANPLAEIN-----FGSKGARA 681
Qy 575 D-IILDAQAISFKGAASRLFLANVS 597
Db 682 DTVLNV-----GEGVNLVATNIT 699

RESULT 11

Q93QW9 PRELIMINARY; PRT; 1654 AA.
AC Q93QW9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE OmpB.
GN OMPB.
OS Rickettsia felis (Rickettsia azadi).
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=42862;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21183306; PubMed=11289671;
RA Moron C.G., Bouyer D.H., Yu X.J., Foil L.D., Crocquet-Valdes P.,
RA Walker D.H.;

Qy 538 INAA-----GTVIANDVSGTVAQINIQNNKIFVINAKNA 572
| : : | : : | : : | : : | : : |
Db 599 IGSSKTALNGGNVAINELVIGNNGSVQFAHNTYLTITSTNA 639

PRT; 1616 AA.

Q9KKA5;
 AC 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE OMPB (Fragment).
 GN OMPB.
 OS Rickettsia rhipicephali.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsieae; Rickettsia.
 OX NCBI_TaxID=33992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3-7-6;
 RX MEDLINE=20393643; PubMed=10939649;
 RA Roux V., Raoult D.;
 RT "Phylogenetic analysis of members of the genus Rickettsia using the
 gene coding the outer-membrane protein rOMP (ompB).";
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
 DR EMBL; AF123719; AAF34122.1; -
 DR InterPro: IPR003858; rOMP_rOMP.
 DR Pfam: PF02708; rOMP_rOMP; 1.
 FT NON_TER 1616
 SQ SEQUENCE 1616 AA; 163271 MW; 42E0489239F15EB4 CRC64;

Query Match 12.0%; Score 363.5; DB 2; Length 1616;
 Best Local Similarity 24.5%; Pred. No. 11e-06;
 Matches 184; Conservative 89; Mismatches 240; Indels 239; Gaps 38;

Qy 23 TTSTAAILMTGSG---VLGAARTVTAGAEALAA-----GTNIGPGAGAFVAGSTLQYTGAF 75
 Db 10 TTRAAATTVGAGFDQTVAPANAVALNAVITANANNINLTPAGSF-----NCLF 61
 Qy 76 TVTDADSVRALDLNFAAGLFVTDGSLGSLGVVDG-ANKLAVNDDGLTLTLTG- 133
 Db 62 LDT-----ANNLAV-----TVSADTLGFTITNAANOQNSFDMLGAGKTLITGGI 108
 Qy 134 TAAYCA-----NPALLFQGGQAAANTYATLGNITLGGANAGL-----TIASDPDL 180
 Db 109 TNAQAAYTKNAQNVQAVGNGAALANDLSGVGTIDWGAADSTLVFNLANPTTOKAPLV 168
 Qy 181 GPIPL-----AGNIDGGGIITDNTDAANGTIGNTPA 213
 Db 169 GNNTVIANGVNGTLNVTNGFIQVSDKSFATVKAINDDQGGFIENTDA-----TVGN--A 221
 Qy 214 AQSISGASTLSLGGAVIKATTKLTNAAPVLTLTNANAVLTG-AVDNTTGGDDVGVNLN 272
 Db 222 LNIQAGGATINFG-----TDGTGRVLVLSKNGAATDFNVYTSGLGNLKGIIIFN 271
 Qy 273 --GALSQVTGNIGNTSLATISVGAG-----TATLGG-----AVIKAT--- 308
 Db 272 TVAVAGQLIANAGPANAVIGTNNCAGRAAGFVVSVDNGNAATGAGVYKDMVIOQSANAG 331
 Qy 309 -----TTKLTNAASVLTLTNALVTGAVD-----NTTG 335
 Db 332 GQVNFHIVDVGTDGTTAFKTAASKVAITONSFGTDFGNLAAQITVPDPTMTLTGNFTG 391
 Qy 336 -----GDNVGVNLSG-----ALSQVTGNIGNTSLATINI-GAGVATLDGAVIKATTKL 385
 Db 392 DASNPNGTAVITPAANGTILASADANAVTNNTIAESGIGVQLSGT--HTAELRL 449
 Qy 386 TDDASVLIFNPVVVTGAIDNT-----GNANKGVVIFTGASTVTDNIGN-----TAVLAESV 438
 Db 450 GNAGSVFKLADGTVINGKVNQAVVGGALAAAGAITLDGSAITIGDINGGGAALQGITL 509
 Qy 439 -----GAGLQIQGVV--KAN--AINLTD-----NAS 462
 Db 510 ANDATKTLTLGGANITGAVGGTIDFOANGGTIKLTSTQNNILVDFDITDQTVGVVDAS 569
 Qy 463 VVFTFGDSTVTGSGT-----GTELFATVNICAGITLRRAGSLAANNIDFGAASNLFNPGA 518
 Db 570 SLTNAQTLTISGIGTIGANKKILGQENKILGQENKILGQENKILGQENKILGQENKILGQ 626
 Qy 519 GKYNLIGT-----IAN-----GNATNLINAACTVIANDVYSIGTVAQIQNNKIFV 566

Db 627 HNNYLITTTNAADQGIIFNPVVNNNTTL---AAGTNLGS--AANPLAELN----- 673
 Qy 567 INAKNADVD-ILDAQAIKFKAASRLFLANVS 597
 Db 674 FSGKAHADTILNV-----GEGVNLATNIT 699

RESULT 14
 Q8XVI3
 ID Q8XVI3 PRELIMINARY; PRT; 3165 AA.
 AC Q8XVI3;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Probable hemagglutinin-related protein.
 GN RSC1775 OR RS04371
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 CC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 CC Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Saranoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 DR EMBL; AL646066; CAD15477.1; -
 DR InterPro: IPR000267; Asp/Glutamase.
 DR PROSITE: PS00144; ASN_GLN_ASE_1; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 3165 AA; 316300 MW; ABD4C7FDDF4696F CRC64;

Query Match 12.0%; Score 361.5; DB 16; Length 3165;
 Best Local Similarity 25.7%; Pred. No. 2.5e-06;
 Matches 181; Conservative 72; Mismatches 273; Indels 177; Gaps 36;

Qy 19 TALETTTAAIMLTGSGVL-----GAARTVTADGAELAGTNGPGAGFVAGSTLOYT 72
 Db 262 TAIAGTGPAPQVAIDVGOLGSMYGGGATRLI---GTEQGLGVNIGNLAALTGRDLN 318
 Qy 73 GAFTVTD-----ADVSVRALDLNFAAGLFVTDGSL-GSVVDG---GANKLAVN 120
 Db 319 GDVITPTGRVQSNADLAIAAPNVN--QGVISTPGVNSISGSTANTGVSVAAGN-VAI- 374
 Qy 121 IDGLTLTLTGTTAAYGANPALLFQGGQAAANTYATL-CNITILGGANAGLTAS---- 175
 Db 375 --AGPQITNTGTIGAGVDANGVT-QAGSVLNAAGTVRNGSLLAQDIGVSGSIDTG 431
 Qy 176 --DPDVLGPIPL--AGNIDGGGIITDNTDAI--NGTIGNTPAAQISIGASTLSLGGAV 229
 Db 432 NGSVNARGAVYTLAAGDSVSRGAASVANSVAIQAGTLDNA-----AGSLMSTTGMQ 483
 Qy 230 IKATTTKLTNAAPVLTLTNANAVLTGAVDNTTG--GDDVGLNLN--GALSQV----- 278
 Db 484 VGA--QRVNVQGGLLGAGVDVATAGSDVDSAGTIGSQTGGLNVNSAGATANAGGKLVA 541
 Qy 279 -----TGNIGNTNSLATISVGAGTATL---GGAVIKATTT-----KLTNAASVLTL 321
 Db 542 QEASLTGTLGNGQGTIVSVRNLSINTGTGGIDNTGTGTVSAAGTAAIGAGALVNRGGTLAA 601
 Qy 322 TN--AVLTGAVDNTTG-----DNVG--VVNLSG----- 346
 Db 602 VGDVALKVGRLDNTAGALGQTAGLKLDSTGDVYNAGGKLVAQAQASIAAASINSOGGTI 661
 Qy 347 -----ALSQVTGNIGNTNSLATINIGAGVATLDGAVIKATTTKLTDDASVLIFNPVV- 400

```

Db 662 SARNLSLNTGTGSDVNTNGTVSARAAATIA-----AGNLVNRGGTLVAVGDVTAH 711
Qy 401 TGAIDNTGNA---NKGVVIFTCASTVIDNIGNTAVLA-EVSVGAGLLQIQGVVKANAIN 456
Db 712 VGSLDNTGAGFGSGGLNVTISAGAI-DNAGGKLVAAQADATLNAASLGNQGGTISARNLS 770
Qy 457 LTDNASVVFTFGDSTVTVTSGISGTELFATVNIAGITLRAGGSAA-----NNID 505
Db 771 LN-----TGTGAIDNTFKTVSAGATATVDAGSLINQGGTLVAVADVQAVNGRLDNTG 822
Qy 506 FGAASNFENGPAGKYNLIGTIANGNN---ATLINAAGTVIANDVSGTVAQINIQNN 562
Db 823 GGLGSGQ-----SGGLNVTISAGAIIDNAGGKLVAAQADASLTGTSLGNOA-GTVAGRNV--- 872
Qy 563 KIFVINAKNADVDILD-----AQAISEKGAASRLFLANVSLQ 599
Db 873 -----TVNTGTGALDNTGGTAIAAAGLDATAGALTANANGVMQ 909

RESULT 15
Q9KKA8
ID Q9KKA8 PRELIMINARY; PRT: 1615 AA.
AC Q9KKA8:
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE OmpB (Fragment).
CN OMPB.
OS Rickettsia montana.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=33991;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M/5-6;
RX MEDLINE=20393643; PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
DR EMBL; AF123716; AAF34119.1;
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
FT NON_TER 1615
SQ SEQUENCE 1615 AA; 162788 MW; 932812728425FB73 CRC64;

Query Match 11.9%; Score 361; DB 2; Length 1615;
Best local Similarity 25.3%; Pred. No. 1.4e-06;
Matches 188; Conservative 87; Mismatches 249; Indels 218; Gaps 40;

Qy 23 TTSTAAILMTGSGVLGAARTVTADGAELAA-----GTNIGPGAGAFVAGSTLOYTGAF 75
Db 10 TTNGRATTVDGAGFDQTAQVNAAPNAVITANANGINFNTPAGSF-----NGLF 61
Qy 76 TVTDADVSRALDLNNFAAGLSVTDISLSVVD-TGGANKLAVNIDDLTLTLTGTO- 133
Db 62 LCT-----ANNLAV---TVSADTTLGFVTNVNNGSNFLTLGAGKTLTITGQGI 108
Qy 134 TAAYGA-----NPALLFGGGAANNVTYALGNITLGCANAGL--TASDDPVLGPITL 185
Db 109 TNAQAAVTNNAAQNVVAAQFNGGVAVANNDLGSGAIDFGAAASTLVFNLAEPPTQKAPLIL 168
Qy 186 AGN-----IDGGGIITDN-----TDAAIN-----GTIGNTPAAQISIGASTLS 224
Db 169 GGNVAVIANGVNTLVNTNGFIQVSKSFATVKAINIGDQGGFI FNTNVAA-----GGNALN 224
Qy 225 L--GGAVIKATTKLTNAAPVLTLTNANAVLTG-AVDNNTGGDDVGVNLN--GALSQVT 279
Db 225 LQVGGATINFNGTGDGTGR---LVLLSKNGAATDFNVNVTSLGGLKNGIIEFNTVAVAGQLI 281
Qy 280 GNIGNTNSLA-----TISVGAG-TATLGG-----AVIKAT----- 308
Db 282 ANAGPANAVIGTNDAGRAAGFWVSYGNGNAATITGVYAKDMVITQSANAGGVNFGHIV 341

```

```

Qy 309 -----TTKLTNAASVLTLTNAVLTGAVD-----NTTGG-----DNV 339
Db 342 DVGDTGTTAFATAASKVAITONSFGTTDFGNLAAQIIVPDTMTLTGNTGNASNPETA 401
Qy 340 GVVNLSG-----ALSOVTGNIGNTNSLATINI-GAGVATLDGAVIKATTTKLTDDASVLIF 394
Db 402 GVIITFAANGTLASASADANVAVTNNITAEASGVGVQLSGT--HTAELRLGNAGSVFKL 459
Qy 395 TNPVVVTGAIDNT-----GNANKGVVIFTGASTVTDNIGN---TAVLAESV----- 438
Db 460 ADGTVINGKVNQTA VVVGALAAAGAITLDGSATITGIDGNGGGAALOGITLANDATKLT 519
Qy 439 --GAGLL-----QIQGGVVKAN-----AINLTD-----NASVVTFTGDS 470
Db 520 LGGANIIGANGGTINFGANGGTIKLTSTONNILLVDFDLAIN-TDQTGVVDASSLINAQTL 578
Qy 471 TVTGSIG-----GTFLFATVNIAGITLRAGGSLAANNIDFGAASNLEF-----NG 516
Db 579 TISGTIGTVGANNTKLGQFNVGSSKLTALNGGNVAINELVIGNNGSVQFAHDTYLTITKTTN 638
Qy 517 PAGKNYNLIGTIANGNNATLINNAAGTVIANDVSGTVAQINIQNNKIFVINAKNADVD- 575
Db 639 AAGQGIIFNPVNVN--NNTTL---AAGTNLGS--AANPLAEIN-----FGSKGAHADT 684
Qy 576 ILDAQAISFKGAASRLFLANVS 597
Db 685 ILNV-----GKGVNLYATNIT 700

```

Search completed: April 14, 2003, 16:11:05
Job time : 102 secs

THIS PAGE BLANK (USPTO)